

**Geospace Model Evaluation  
to Support Model Transition to Operations.  
Phase II Report:  
Regional *K*-Index  
(Version 2013/08/31)**

*Community Coordinated Modeling Center*

A. Glocer, L. Rastätter, A. Pulkkinen, M. Kuznetsova

*CCMC, NASA Goddard Space Flight Center*

Submitted: 2013/08/31

Updated: 2013/09/06

---

**Abstract.**

We present the latest result of a community-wide space weather model validation effort coordinated among the Community Coordinated Modeling Center (CCMC), NOAA Space Weather Prediction Center (SWPC), model developers, and the broader science community. Validation of geospace models is a critical activity for both building confidence in the science results produced by the models and in assessing the suitability of the models for transition to operations. Indeed, a primary motivation of this work is supporting NOAA SWPC's effort to select a model or models to be transitioned into operations. Our validation efforts focus on the ability of the models to reproduce a location specific index of geomagnetic disturbance, the local k-index. Six geomagnetic events representing a range of events and six geomagnetic observatories representing mid- and high-latitude locations is considered in our analysis. Model performance is evaluated quantitatively by the use of contingency tables, skill scores, and distribution metrics. We consider model performance on an event-by-event basis, aggregated over events, at specific station locations, and separated into high- and mid-latitude domains. The summary of results are presented in this report, and an online interface built at CCMC is available for detailed time series analyses.

## 1. Introduction

Forecasting the level of geomagnetic disturbance on the ground is critical step in mitigating the potentially severe impact of geomagnetically induced currents (GICs) [e.g., *Boteler et al.*, 1998; *Pirjola*, 2005; *North American Electric Reliability Corporation*, 2012; *National Research Council*, 2008]. The science community has responded with both first principles and empirical models capable of forecasting these potentially hazardous disturbances. Before such models can be transitioned in to an operational setting, however, a comprehensive model validation effort is required. The Community Coordinated Modeling Center (CCMC), NOAA Space Weather Prediction Center (SWPC), model developers, and the broader science community have joined together to carry out this important validation effort. This report represents the latest model validation findings in support of geospace model transition to operations.

This Phase II report builds on the prior studies of geospace model validation [*Pulkkinen et al.*, 2010, 2011; *Rastätter et al.*, 2011], and in particular on our Phase I report based on [*Pulkkinen et al.*, 2013] and *Rastätter et al.* [2013]. The Phase I report focused on the ability of models to reproduce  $dB/dt$  (i.e. rapid fluctuation of the ground magnetic field) at specific magnetometer locations. We encourage the reader to refer back to that work as this study is a direct follow on to that effort. As the work on the Phase I report was coming to completion, work was initiated on Phase II, to consider the ability of models to reproduce a local index of geomagnetic disturbance.

The  $Kp$  index is a commonly used global measure of geomagnetic disturbances. It is a measure on a scale of 0-9, based on a scaled range of  $\Delta B$ , of the average level of

disturbance. Local predictions of  $K$ , however, may differ significantly from the global average. The interest in predicting potential GICs and geomagnetic disturbances on a local level, and the convenience of an index of activity instead of a raw prediction, provides the motivation for the present report.

The structure of the report is as follows. In Section 2 we will describe the setting used in the validation effort. Section 3 details the metrics used in the quantification of the model performance and in Section 4 each participating model is summarized. The main results of the validation effort are reported in Section 5. Finally, Section 6 provides a brief discussion of our findings.

## 2. Validation setting

As noted in the previous section, the present work builds on the validation study presented in the Phase I report. To avoid repeating the very complete description of the validation setting provided previously, we will only provide an overview here as well as new features particular to the current study. An interested reader may refer to the attached Phase I report and to [Pulkkinen *et al.*, 2013] for further details.

Six events were chosen for the study consisting of the four events from the earlier GEM Challenges [Pulkkinen *et al.*, 2010, 2011; Rastätter *et al.*, 2011] as well as two “surprise events” not communicated to the modelers prior to their delivering their models to CCMC for evaluation. The surprise events were selected jointly by CCMC and NOAA SWPC scientists. The event list is given in Table 1.

Six stations representing the high-latitude and mid-latitude locations were selected out of the original twelve GEM Challenge stations. The high-latitude stations are PBQ/SNK, ABK and YKC and mid-latitude stations WNG, NEW, OTT (see Table 1 and Fig. 1). In

the case of the global MHD models, the magnetic field variations at each magnetometer location was computed by a Biot-Savart integral over the entire domain. The integration includes all currents in the magnetosphere, as well as the field-aligned currents in the gap region between the MHD model’s inner boundary and the ionosphere, and the high-latitude ionospheric currents. The CCMC tool used for the integration is described in detail in the Phase I report and in the paper by *Rastätter et al.* [2013] and is applied to each of the Global MHD models used in the study. The two empirical models (see Table 3) provided direct predictions of the magnetic field at the used station locations. All model runs and ground magnetic field calculations (with the exception of Wing Kp) were carried out at CCMC.

For each event in Table 1, the model performance was evaluated by comparing the observed vs predicted local  $K$ -values at the specific magnetometer locations listed above. Throughout the paper  $K$  is calculated in the following way. First we find the maximum “Range” of  $\Delta B$  in the two horizontal directions.

$$Range = \max [(\Delta B_{x,max} - \Delta B_{x,min}), (\Delta B_{y,max} - \Delta B_{y,min})] \quad (1)$$

over a three-hour window sliding by 15 minutes. where  $B_{x,max}, B_{x,min}, B_{y,max}$ , and  $B_{y,min}$  indicate the max and min values in the window of the two horizontal components of the magnetic field (**geomagnetic dipole coordinates**). The Range is then scaled with a station specific scaling factor. Scaling factors for stations used in this validation study are given in Table 2.  $K$  is then found from the scaled range using a lookup table. The same approach was used for both models and observations. Again following the earlier GEM Challenges, 12 geomagnetic observatories (magnetometer stations) listed in Table 2 and shown in Fig. 1 were selected based on the global spatial and temporal coverage. Three

high-latitude and three mid-latitude stations (the same as for the Phase I report) were included in the present study (bold in Table 2). Station PBQ was discontinued November 2007 and replaced by station SNK. Consequently, for events 5 and 6 station SNK was used in place of PBQ. We use the results from the model and observations from the Phase I V.20130831 report to get the time series used to calculate  $K$  in this study. No new models runs or data processing was carried out to get the time series from which we calculate the local  $K$  value.

Note that 5\_WEIMER outputs used in the V20130419 of the Phase I report and in the *Pulkkinen et al.* [2013] paper were replaced with the 6\_WEIMER outputs in the V.20130831 reports. The 6\_WEIMER outputs generated for V.20130831 reports (dB/dt and Regional-K) reflect the recent corrections that were found necessary in running the Weimer model: 1) the driver routine now properly feeds in  $B_y$  from the solar wind; 2) applies a 17-minute delay to the solar wind input; 3) rotates model output into magnetic dipole coordinates. Any updates to time series introduced during the revision of the Phase I report are taken into account in this study. The details and timing of all simulation runs updates for all models introduced since the initiation of the study in January 2011 are summarized in the "Milestones of model deliveries and run executions" section of the V20130419 of the Phase I report.

### 3. Selected metrics

We use the same metrics as from the earlier study of *Pulkkinen et al.* [2013] with a few modifications. In particular, the model validation is largely built on event-based analyses. An event is defined here as follows: the value of the parameter,  $K$ , at a given time of interest exceeds an event threshold  $K_{thres}$ . The windows for calculating  $K$  are

three-hours in size, centered around the time of interest, and moved in 15 minute steps over the time series in overlapping 3-hour segments. Events for given time and  $K_{thres}$  are recorded for both the measured and the modeled  $K$ . By comparing threshold crossing for both observed and modeled time series one can then build a four-element matrix known as contingency table. The table reports the number of correct hits, false alarms, missed events and correct no events [e.g., *Lopez et al.*, 2007]. In this work the thresholds for  $K$  were chosen to be high,  $K = 6$ , and very high,  $K = 8$ . **The selected thresholds are chosen with the idea that higher  $K$  values representing stronger events are of more interest for space weather applications.**

The elements of the contingency table contain the number of correctly predicted threshold crossings  $H$  (hits), the number of false alarms  $F$ , the number of missed crossings  $M$  and the number of correctly predicted no crossings  $N$ . The set  $\{H, F, M, N\}$  can be used to compute a number of different metrics quantifying the performance of individual models. In this study three metrics, proposed by NOAA SWPC, were selected for use in the final analyses. The selected metrics are Probability of Detection (POD), Probability of False Detection (POFD) and Heidke Skill Score (HSS). For interest, we also include the Critical Success Index (CSI) as an additional skill score. CSI have not been used as a base for model ranking. We describe each metric more in detail in the following subsections.

In addition to the event tables and skill scores, we also consider a so-called distribution metric. In this metric, we consider the distribution of model predictions when the observations are a particular value of  $K = K_0$ . A model that performs well in this metric would show a distribution peaked around  $K_0$  with very little spread in the distribution. A model with significant random error would exhibit broadening of the distribution. A model with

systematic error would have the distribution shifted so the peak is above or below  $K_0$ . In this study we consider the distribution metric for three values of  $K = 4, 6, 8$ , and qualitatively compare the results to see the relative presence of random and systematic error in model results.

### 3.1. Probability of Detection

POD is defined for the set  $\{H, F, M, N\}$  as

$$POD = \frac{H}{H + M} \quad (2)$$

The metric measures the fraction of observed threshold crossings which were correctly forecast. It ranges from 0 to 1 with 1 being a perfect score. Since a model providing artificially large signal amplitudes will tend to generate large  $H$  and large POD the metric should be used in conjunction with POFD defined below.

### 3.2. Probability of False Detection

POFD is defined for the set  $\{H, F, M, N\}$  as

$$POFD = \frac{F}{F + N} \quad (3)$$

The metric measures the probability of an incorrectly forecasted threshold crossing when no such crossing occurs. POFD ranges from 0 to 1, with 0 being a perfect score. Similar to POD, a model predicting artificially low signal amplitudes will provide low  $F$  and small POFD and thus the metric should be used in conjunction with POD.

### 3.3. Heidke Skill Score

HSS is defined for the set  $\{H, F, M, N\}$  as

$$HSS = \frac{2(HN - MF)}{(H + M)(M + N) + (H + F)(F + N)} \quad (4)$$

The metric measures the fraction of correctly predicted threshold crossings after eliminating those predictions that would be correct purely by random chance. It ranges from negative infinity to 1. Negative values indicate that random forecast is better than the model prediction, 0 indicates no skill (as good as random) and 1 indicates a perfect score.

### 3.4. Critical Success Index

CSI is defined for the set  $\{H, F, M, N\}$  as

$$CSI = \frac{H}{H + M + F} \quad (5)$$

The metric ranges from 0 (no skill) to 1 (perfect skill) and accounts both false alarms and missed events. We include it here only as an alternative, but less sophisticated way, of evaluating skill than the HSS.

## 4. Models

We include the same five models used in the Phase I evaluation study. These included empirical models by D. Weimer (Virginia Polytechnic Institute) and R. Weigel (George Mason University) and major US global magnetohydrodynamic (MHD) models from University of Michigan, Center for Integrated Space Weather Modeling (CISM) and University of New Hampshire. In addition to these models, we also include the WingKp model of Global Kp prediction. This last model was added in order to determine the “value

added” of models that can predict local  $K$  values, compared with a model currently used to predict magnetic disturbance levels (and assuming that prediction applies everywhere).

All models that participated the validation effort were delivered to CCMC. CCMC had extensive communications with the model developers to guarantee correct installation and to ensure the usage of appropriate settings for each model. Based on a variety of tests such as code robustness carried out at CCMC, model developers provided revisions to the model settings. The final selection of all model settings was accomplished by mid-August 2011. To allow for simulations in a realistic real-time computational environment, it was required that settings for all models were such that the simulations would run not slower than twice the modeled physical time on 64 Beowulf cluster processors. Detailed model descriptions and milestones of model deliveries and run executions are presented in the attached Phase I V20130831 report. **All simulations, except for WingKp, were performed at CCMC using the same computational architecture.**

The WingKp model was never delivered to CCMC and was therefore run at AFRL. Additionally, AFRL was not able to provide results for event 3 which was outside their run window. The WingKp model also used a different representation of the solar wind input, and occasionally was not able to supply a prediction due to missing data. Such predictions show up as a no data flag ( $K=-1$ ) in the online plotting, and are excluded from our metrics analysis. WingKp was handled differently because it only was used in this study to compare the local prediction of  $K$  by the models under evaluation with a  $Kp$  prediction that is currently available to SWPC forecasters.

Table 3 summarizes some of the key features of each individual model. A version of the Weimer model and all global MHD models discussed in this work are available at CCMC

for runs-on-request. For further description of the models used, please refer back to the Phase I of the geospace validation.

## 5. Results

Figure 2 shows an example time series of the observed vs modeled  $K$  for the event 2 (Table 1). Each model is shown in a separate panel (red line) together with the observations (black line). We chose a random mid-latitude station for this demonstration. All data are viewable via CCMC's online visualization interface accessible at [http://ccmc.gsfc.nasa.gov/challenges/dBdt/metrics\\_results.php](http://ccmc.gsfc.nasa.gov/challenges/dBdt/metrics_results.php). We will quantify this capability to capture the events using metrics discussed in Section 3.

The final metrics-based analyses were carried out for each individual model using events and stations described in Section 2 and the corresponding contingency tables with elements  $\{H, F, M, N\}$  were generated for each model. In the following subsections we will present the results broken out in several ways.

### 5.1. All Stations and All Events

Here we will report the results summed over all events recorded for all stations, i.e. the results are integrated over high-latitude (PBQ/SNK, ABK, YKC) and mid-latitude stations (WNG, NEW, OTT). Figure 3 and tables 4 and 5 show the performance.

### 5.2. High-Latitude Stations and All Events

Here we will report the results summed over all events over high-latitude (PBQ/SNK, ABK, YKC) stations. Figure 4 and tables 6 and 7 show the performance.

### 5.3. Mid-Latitude Stations and All Events

Here we will report the results summed over all events over mid-latitude (WNG, NEW, OTT) stations. Figure 5 and tables 8 and 9 show the performance.

### 5.4. Separating First Four Events from Last Two (surprise) Events

It is valuable to further separate the model performance observed in the first four events from the last two events. The reasoning is simple: The first four events were known to developers before submitting models for the competition while the last two events were so-called “surprise” events that were not known to the model developers. The “surprise” events exclude an unlikely possibility that model developer may have decided to optimize their model for the known events.. We therefore present the results for the first four and last two events separately. However, we only consider a threshold of 6 in this case in order to maintain reasonable statistics for the lower number of events.

Figure 6 and tables 10 and 11 show the performance for the first four events. Figure 7 and tables 12 and 13 show the performance for the last two events.

### 5.5. A Cursory Examination of Lower Thresholds

This study is primarily interested in strong and very strong events, which is why we consider thresholds of  $K=6$  and 8. Now, however, we will test the sensitivity of our results to threshold by considering the results for a threshold of  $K=4$ . Here we will report the results summed over all events over all high- and mid-latitude stations. Figure 8 show the performance.

## 5.6. Distribution Metric

As described in Section 3, we also incorporate a “distribution” metric. The concept behind this metric is straight forward: We look at the distribution of model predictions at a particular stations for a observed  $K$  at that same station. A model that performs in an ideal way, will have its distribution of model predictions peaked sharply at the observed  $K$ . The presence of random error in the model prediction appears as a broadening of the distribution. Systematic error manifests itself as a shift in the peak of the distribution relative to the observed  $K$ .

Although we do not employ a mathematically rigorous analysis of the model performance in the distribution metric, a great deal can still be learned by visual inspection of the distributions. For instance, a peak shifted to the left represents and systematic under-prediction while a peak shifted to the right represents a systematic over prediction. When taken conjunction with the contingency tables and skill scores the results can be quite illuminating. A model that has a high-probability of false detection, for instance, could have those false detections as a result of systematic error causing the model to consistently predict higher values, random error causing the model to false detections randomly, or a combination of both. The contingency tables alone cannot pinpoint the type of error, but including the distribution metric can provide insight. We consider the results station-by-station to gain a more granular picture of model performance and understand if some stations may be contributing more strongly (positively or negatively) to the skills scores calculated from contingency tables. One important factor to keep in mind is that the number of events decreases for  $K = 8$  and may be very small when considering the distribution on a station-by-station basis.

The results for the distribution metric are presented in Figures 9 through 22. Each figure presents results for  $K=4$  (left column),  $K=6$  (middle column), or  $K=8$  (right column). Additionally, each row presents results for a different magnetometer station. Every model has two figures associated with it; one for high-latitude stations and one for low latitude stations. In the following paragraphs we will describe the results for each model.

Figures 9 and 10 present the high- and mid-latitude results for the WingKp Model. The mid-latitude results for an observed  $K=4$  and  $K=6$  demonstrate a very large spread indicating significant random error. For  $K=8$ , the results are more peaked at the correct value of  $k$  although some random error is still visible. The results are similar for high-latitude which is consistent with the event based analysis. The results for station PBQ are particularly good with peaks at the correct values of  $K$ , albeit with some spread. However, the results for stations YKC and ABK exhibit significant random error for all values of  $K$ . As WingKp produces a single global prediction of  $K_p$ , and we are using that prediction for local  $K$  predictions, some error is to be expected. From this type of analysis we can see that the error is more random in nature.

Figures 11 and 12 present the high- and mid-latitude results for the 9\_SWMF Model. For mid-latitude stations the distributions are typically peaked at or near the correct values of  $K$ . Some moderate spread in the distributions are present indicating the presence of some random error. The same largely holds true for high-latitude results with the spreading a bit more pronounced. Also a slight systematic shift towards under-prediction is seen when the observed  $K=8$ . This is consistent with the trend seen in the event studies that performance for 9\_SWMF was stronger for mid-latitude compared to high-latitude. It is also consistent with the finding from the event table that 9\_SWMF has higher skill

for threshold of  $K=8$  (compared to  $K=6$ ) for mid-latitude, but the reverse is true for high-latitude .

Figures 13 and 14 present the high- and mid-latitude results for the 9a\_SWMF Model. The findings for 9a\_SWMF are largely similar to those for 9\_SWMF. That is expected as they are the exact same model, just with different implementations for calculating the magnetic field perturbations on the ground. For 9\_SWMF the magnetic field perturbation on the ground is produced by the CCMC post-processing delta-B tool used for all global magnetosphere models evaluated in this report. For the 9a\_SWMF the magnetic field perturbation is calculated internally to the SWMF model.

Figures 15 and 16 present the high- and mid-latitude results for the 2\_LFM-MIX Model. For both mid- and high-latitude stations, the distribution of model predictions for an observed  $K$  tend to peak below the observed value of  $K$ . This shift in the peak of the distribution relative to the observed  $K$  is indicative of a systematic under-prediction by the model. The 2\_LFM-MIX model was found to have extraordinarily low POFD in the event based analysis which is likely a result of this systematic shift. Some modest evidence of random error is visible in the spreading of the distribution, but it is not enough to result in significant false detections for the  $K$  thresholds considered.

Figures 17 and 18 present the high- and mid-latitude results for the 4\_OPENGGCM Model. One feature that stands out is the large number of occurrences in the model predictions of  $K$  values greater than then observed  $K$ . Sometimes this is a systematic shift in the distribution (e.g., WNG and NEW ,  $K=4$ ), and sometimes it appears to be more random error (e.g., OTT  $K=4$  and NEW  $K=6$ ). Regardless of whether the shift is systematic or random, the high-occurrence of predictions significantly exceeding the

observations, particularly for mid-latitude stations and lower  $K$  values, results in lots of false detections (even if true detections are still plentiful). This finding is consistent with the high-POFD exhibited by 4\_OPENGGCM in the event studies.

Figures 19 and 20 present the high- and mid-latitude results for the 2\_WEIGEL Model. For both mid- and high-latitude stations, and for all choices of observed  $K$ , the distribution of model predictions is peaked below the observations. Such a shift represents a systematic under-prediction of the model. As a result, the model is likely to have a low POFD. These findings are consistent with the event based analysis which demonstrates that the 2\_WEIGEL model has exceedingly low POFD.

Figures 21 and 22 present the high- and mid-latitude results for the 6\_WEIMER Model. For both mid-latitude stations, for observed  $K=4$  and  $K=8$ , the distribution of model predictions is peaked below the observations. For  $K=6$  the distribution of model predictions is peaked right at 6 for 2 out of 3 stations. For high-latitude stations for all observed values of  $K$  the distribution is seen to be shifted to the left representing a systematic under prediction. This pattern seems consistent with the event based studies when the model showed low POFD (owing apparently to the systematic under-prediction) and the strongest performance among models for mid-latitude stations when the  $K$  threshold is set to 6, but worse performance for higher  $K$  threshold and high-latitude.

In summary, the distribution metric, is quite useful in understanding why the event based metrics came out as they did. We are able to see visually the presence of systematic and random error and found that models with large amounts either can affect the POD and POFD (either positively or negatively).

## 6. Discussion

This work represents the Phase II of the geospace model validation effort, carried out in coordination among the CCMC, NOAA SWPC, modelers and science community. The effort is a continuation of the earlier GEM modeling challenges and the  $dB/dt$  validation study summarized in the Phase I report. The focus of the effort was to evaluate the ability of geospace models in predicting the local  $K$  index and moreover, as for  $dB/dt$ , evaluate the potential value added of a local prediction over the a global prediction.

We considered two types of metrics in evaluating the model  $K$  prediction: skills scores calculated from event based contingency tables and a distribution metric. The skills scores (POD, POFD and HSS) from event based contingency tables for different  $K$  thresholds were the primary metric we used to rank the models. In particular, the HSS which reflects how much better a model is compared to random chance. The derived contingency tables were compiled in several different ways. We grouped all the stations and events together, we separated high-latitude stations and mid-latitude stations for all events, and we separated events into those known to the model developer ahead of time (first four events) and the surprise events selected after models were delivered to CCMC for evaluation (last two events). These different groupings allow us to draw more detailed conclusions about model performance and suitability of models for forecasting  $K$  values at mid-latitude vs high-latitude and strong events vs very strong events. The distribution metric was an additional tool used more to gain insight into the model performance(i.e., how prevalent was random error, systematic error, ...).

In terms of actual model performance, the 9\_SWMF and 9a\_SWMF models were consistently strong performers in all the metrics almost always ranking near the top in all

categories. The model had relatively high-POD and low POFD resulting in HSS that was always among the best. The distribution metric revealed the presence of a moderate amount of random error and limited systematic error. We reiterate that 9\_SWMF and 9a\_SWMF are actually the same model except that the magnetic field perturbation on the ground is calculated by CCMC (for 9\_SWMF) and internally to the SWMF model (for 9a\_SWMF). Hence similar performance is expected.

The 2\_LFM-MIX model typically had lower performance compared to other models as measured by HSS. The exception was the last two events for mid-latitude where the model performance was in the middle of the pack. The model typically exhibited lower POD and POFD. The distribution metric shows a clear tendency of this model to under-predict  $K$  and that likely results in the lower POD, POFD, and HSS. We note that these results are consistent with the earlier  $dB/dt$  study in which the 2\_LFM-MIX model performed worse for larger thresholds of magnetic perturbation. It is possible that the model would perform better if we used lower  $K$  thresholds for calculating the contingency tables, just as the model did better in the  $dB/dt$  study for lower thresholds. However, the present study is focused primarily on the model's ability to detect strong and very strong disturbances, not small or moderate disturbances. Our cursory examination of a lower threshold of  $K=4$  did not result in a significant change in the ordering of models by performance (although HSS did increase).

The 6\_WEIMER statistical model performed exceptionally well for mid-latitudes for a threshold of  $K=6$ , the top performer in this category. The model performance decreased significantly for mid-latitudes with a threshold of  $K=8$ , but the performance was still

strong. In contrast to mid-latitudes the model performance dropped significantly at high-latitude for both  $K$  thresholds.

The 4\_OPENGGCM model had mixed performance. It generally had very good POD, but had a consistently elevated POFD. As seen in the distribution metric, the model had a tendency to overpredict, that leads to the high POD and high POFD. As a result, sometimes the model has a good HSS and sometimes not so good depending on how strongly the POD outweigh the POFD. Significant random and systematic error was likely the cause of the the higher POFD. Regardless of the cause, and overall result on the HSS, an elevated POFD is a concern that needs to be considered in an operational setting. The model did perform better in the last two events compared to the first four.

The 2\_WEIGEL model was never the top performing model, but it was also never the worst performing model as measured by HSS. In the distribution metric, the model was seen to typically under-predict the observations and as a result have an exceedingly low POFD with still a reasonable POD.

One of the key questions we tried to answer in the study is: “To what degree do geospace models predict a local  $K$  value better than assuming the current prediction of global  $K_p$  applies everywhere.” To answer this question we included the WingKp model, which is currently used by SWPC as one input to short-term  $K_p$  predictions, in our analysis, and took the global  $K_p$  prediction as a prediction of local  $K$ . Using global  $K_p$  predictions from the WingKp model was never a top performing approach for local  $K$  predictions as measured by HSS. Interestingly, the model used in this way was also often not the lowest performing model, indicating that using the WingKp prediction of global  $K_p$  (as a local  $K$  prediction) would actually exhibit higher skill than using the local  $K$  predicted

by some models. The POFD was typically elevated, however, compared to other models. An elevated POFD raises concerns of using the global Kp prediction from WingKp for local forecasts of  $K$ , but it also demonstrates the potential value of a local  $K$  forecast. All local  $K$  forecasts (except for 4\_OPENGGCM) consistently had much lower POFD.

All the models we examined had positive HSS demonstrating better prediction skill than random chance. Moreover, we found most of our results to be consistent with the  $dB/dt$  study of *Pulkkinen et al.* [2013]. When considering all events, POD of around 70% are possible for the top performing models for mid-latitude stations, even with a  $K$  threshold of 8. For high-latitude stations, the POD possible for top performing models drops to around 50%. In either case, the POFD for most models are exceedingly low for the thresholds considered. Whether this performance is sufficient for current space weather prediction needs, or if further improvement is required is not a question addressed in this study. Nevertheless, these results represent what is currently possible with the current state-of-the-art.

### **Acknowledgments.**

The results presented rely on data collected at magnetic observatories. We thank the national institutes that support them and INTERMAGNET for promoting high-standards of magnetic observatory practice ([www.intermagnet.org](http://www.intermagnet.org)). We also would like to thank the entire team at NOAA SWPC, led by Howard Singer in this effort, for their valuable contributions in defining the metrics, in particular Chris Balch for providing the algorithm for calculating  $K$ . We are also grateful for the efforts of James McCoullough (AFRL) for providing the Wing Kp results in form that it could be included in this study, and to Simon Wing for use of results from his model. CCMC staff is acknowledged for their support

throughout. Finally, none of this work could have been conducted without the support, input, and time of the model developers who contributed models to this effort. The LFM-MIX model team represented by Michael Wiltberger, the OpenGGCM model team led by Jimmy Raeder, the SWMF model team represented by Gabor Toth and Dan Welling, and empirical modeling efforts led by Daniel Weimer and Bob Weigel were instrumental in making this study happen.

## References

- Baker, D. N., R. S. Weigel, E. J. Rigler, R. L. McPherron, D. Vassiliadis, C. N. Arge, G. L. Siscoe, and H. E. Spence (2004), Sun-to-magnetosphere modeling: CISM forecast model development using linked empirical methods, *Journal of Atmospheric and Solar-Terrestrial Physics*, 66, 1491, doi:10.1016/j.jastp.2004.04.011.
- Boteler, D.H., R.J. Pirjola, and H. Nevanlinna (1998), The Effects of Geomagnetic Disturbances on Electrical Systems at the Earth's Surface, *Adv. Space Res.*, 22, 17-27.
- Buzulukova, N., M.-C. Fok, A. Pulkkinen, M. Kuznetsova, T. E. Moore, A. Glocer, P. C. Brandt, G. Toth, and L. Rastatter (2010), Dynamics of ring current and electric fields in the inner magnetosphere during disturbed periods: CRCM-BATS-R-US coupled model, *J. Geophys. Res.*, 115:A05210.
- Connor, H. J., J. Raeder, and K. H. Trattner (2012), Dynamic modeling of cusp ion structures, *J. Geophys. Res.*, 117, A04203, doi:10.1029/2011JA017203.
- D.L. De Zeeuw, S. Sazykin, R.A. Wolf, T.I. Gombosi, A.J. Ridley, and G. Tóth (2004), Coupling of a global MHD code and an inner magnetosphere model: Initial results, *J. Geophys. Res.*, 109(A12):A12219.

- Dorelli, J. C., M. Hesse, M. M. Kuznetsova, L. Rastaetter, and J. Raeder (2004), A new look at driven magnetic reconnection at the terrestrial subsolar magnetopause, *J. Geophys. Res.*, 109, A12,216, doi: 10.1029/2004JA010,458.
- Emmert, J. T., A. D. Richmond, and D. P. Drob (2010), A computationally compact representation of MagneticApex and QuasiDipole coordinates with smooth base vectors, *J. Geophys. Res.*, 115, A08322, doi:10.1029/2010JA015326.
- Fuller-Rowell, T. J., D. Rees, S. Quegan, R. J. Moffett, M. V. Codrescu, and G. H. Millward (1996), A coupled thermosphere-ionosphere model (CTIM), in *STEP Report*, edited by R. W. Schunk, p. 217, Scientific Committee on Solar Terrestrial Physics (SCOSTEP), NOAA/NGDC, Boulder, Colorado.
- Garcia, K. S., and W. J. Hughes (2007), Finding the Lyon-Fedder-Mobarry magnetopause: A statistical perspective, *J. Geophys. Res.*, 112, 06229, doi:10.1029/2006JA012039.
- Ge, Y. S., J. Raeder, V. Angelopoulos, M. Gilson, and A. Runov (2011), Interaction of dipolarization fronts within multiple bursty bulk flows in global MHD simulations of a substorm on 27 February 2009, *J. Geophys. Res.*, 116, A00I23, doi:10.1029/2010JA015758.
- Gilson, M. L., J. Raeder, E. Donovan, Y. S. Ge, and E. L. Kepko (2012), Global simulation of proton precipitation due to field line curvature during substorms, *J. Geophys. Res.*, 117, A05,216.
- Glocer, A., G. Tóth, M. Fok, and T. Gombosi (2009a), Integration of the radiation belt environment model into the space weather modeling framework, *J. Atmos. Solar-Terr. Phys.*, 71:1653–1663.

- Glocer, A., G. Tóth, T. I. Gombosi, and D. Welling (2009b), Polar Wind Outflow Model (PWOM): Modeling ionospheric outflows and their effect on the magnetosphere, initial results, *J. Geophys. Res.*, 114:A05216.
- Glocer, A., G. Tóth, Y. J. Ma, T. Gombosi, J.-C. Zhang, and L. M. Kistler (2009c), Multi-fluid BATS-R-US: Magnetospheric composition and dynamics during geomagnetic storms, initial results, *J. Geophys. Res.*, 114, A12203, doi:10.1029/2009JA014418.
- Gombosi, T.I., G. Tth, D.L. De Zeeuw, K.C. Hansen, K. Kabin, and K. G. Powell (2002), Semi-relativistic magnetohydrodynamics and physics-based convergence acceleration, *J. Computational Phys.*, 177, 176-205.
- Gombosi, T.I., K. G. Powell, D. L. De Zeeuw, C. R. Clauer, K. C. Hansen, W. B. Manchester, A. J. Ridley, I. I. Roussev, I. V. Sokolov, Q. F. Stout, and G. Tóth (2004), Solution-adaptive magnetohydrodynamics for space plasmas: Sun-to-Earth simulations, *Computing in Science and Engineering*, 06(2):14–35, Mar/Apr.
- Goodrich, C. C., J. G. Lyon, M. Wiltberger, R. E. Lopez, and K. Papadopoulos (1998), An overview of the impact of the January 10-11, 1997 magnetic cloud on the magnetosphere via global MHD simulation, *Geophys. Res. Lett.*, 25(1), 25372540, doi:10.1029/98GL01159.
- Guild, T. B., H. E. Spence, E. L. Kepko, V. Merkin, J. G. Lyon, M. Wiltberger, and C. C. Goodrich (2008a), Geotail and LFM comparisons of plasma sheet climatology: 1. Average values, *J. Geophys. Res.*, 113, 04216, doi:10.1029/2007JA012611.
- Guild, T. B., H. E. Spence, E. L. Kepko, V. Merkin, J. G. Lyon, M. Wiltberger, and C. C. Goodrich (2008b), Geotail and LFM comparisons of plasma sheet climatology: 2. Flow variability, *J. Geophys. Res.*, 113, 04217, doi:10.1029/2007JA012613.

- Haines, G. V. (1985), Spherical cap harmonic analysis, *J. Geophys. Res.*, 90 (B3), 2583.
- Hapgood, M.A. (1992), Space Physics Coordinate Transformations: A User Guide, *Planet. Space Sci.*, Vol. 40, No. 5, 711-717.
- Kennel, C. F., and H. E. Petschek (1966), Limit on stably trapped particle fluxes, *J. Geophys. Res.*, 71, 1.
- Knight, S. (1972), Parallel electric fields, *Planet. Space Sci.*, 21, 741.
- Koren, B. (1993), A robust upwind discretisation method for advection, diffusion and source terms, In C.B. Vreugdenhil and B.Koren, editors, *Numerical Methods for Advection-Diffusion Problems*, page 117, Vieweg, Braunschweig.
- Li, W., J. Raeder, J. Dorelli, M. Oieroset, and T. D. Phan (2005), Plasma sheet formation during long period of northward IMF, *Geophys. Res. Lett.*, 32, L12S08, doi:10.1029/2004GL021,524.
- Li, W., J. Raeder, J. C. Dorelli, M. Thomsen, and B. Lavraud (2008), Solar wind entry into the magnetosphere under northward IMF conditions, *J. Geophys. Res.*, 113, A04, 204.
- Li, W., J. Raeder, M. Oieroset, and T. D. Phan (2009), Cold dense magnetopause boundary layer under northward IMF: Results from THEMIS and MHD simulations, *J. Geophys. Res.*, 114, A00C15, doi:10.1029/2008JA013, 497.
- Li, W., J. Raeder, and D. Knipp (2011), The relationship between dayside local poynting flux enhancement and cusp reconnection, *J. Geophys. Res.*, 116, A08, 301, doi:10.1029/2011JA016,566.
- Lopez, R. E., C. C. Goodrich, M. Wiltberger, K. Papadopoulos, and J. G. Lyon (1998), Simulation of the March 9, 1995 substorm and initial comparison to data, *Geospace Mass*

*and Energy Flow: Results From the International Solar-Terrestrial Physics Program*, 104, 237-245, doi:10.1029/GM104p0237.

Lopez, R. E., S. Hernandez, M. Wiltberger, J. Lyon, and C. Goodrich (2006), Initial Results from the Simulation of the Halloween 2003 Storms, *Advances in Geosciences*, 2, 191.

Lopez, R. E., S. Hernandez, M. Wiltberger, C.-L. Huang, E. L. Kepko, H. Spence, C. C. Goodrich, and J. G. Lyon (2007), Predicting magnetopause crossings at geosynchronous orbit during the Halloween storms, *Space Weather*, 5, S01005, doi:10.1029/2006SW000222.

Lu, G., et al. (2011), Reversed two-cell convection in the northern and southern hemispheres during northward interplanetary magnetic field, *J. Geophys. Res.*, 116, A12,237.

Lyon, J. G., J. A. Fedder, and C. M. Mobarry (2004), The Lyon-Fedder-Mobarry (LFM) global MHD magnetospheric simulation code, *Journal of Atmospheric and Solar-Terrestrial Physics*, 66, 1333, doi:10.1016/j.jastp.2004.03.020.

Lyons, L. R., D. Evans, and R. Lundin (1979), An observed relation between magnetic field aligned electric fields and downward electron energy fluxes in the vicinity of auroral forms, *J. Geophys. Res.*, 84, 457.

Merkin, V. G., and J. G. Lyon (2010), Effects of the low-latitude ionospheric boundary condition on the global magnetosphere, *J. Geophys. Res.*, 115(A), 10202, doi:10.1029/2010JA015461.

Muhlbacher, S., C. J. Farrugia, J. Raeder, H. K. Biernat, and R. B. Torbert (2005), A statistical investigation of dayside erosion showing saturation response, *J. Geophys. Res.*, 110, A11, 207, doi:10.1029/2005JA011,177.

National Research Council (2008), Severe Space Weather Events-Understanding Societal and Economic Impacts: A Workshop Report, *The National Academies Press*, Washington, DC.

North American Electric Reliability Corporation GMD Task Force (2012), 2012 Special Reliability Assessment Interim Report: Effects of Geomagnetic Disturbances on the Bulk Power System, *NERC*, February 2012.

Pembroke, A., F. Toffoletto, S. Sazykin, M. Wiltberger, J. Lyon, V. Merkin, and P. Schmitt (2012), Initial results from a dynamic coupled magnetosphere-ionosphere-ring current model, *J. Geophys. Res.*, 117(A2), doi:10.1029/2011JA016979.

Pirjola, R., Effects of space weather on high-latitude ground systems, *Advances in Space Research*, 36, 2231-2240, 2005.

K.G. Powell, P.L. Roe, T.J. Linde, T.I. Gombosi, and D. L. De Zeeuw (1999), A solution-adaptive upwind scheme for ideal magnetohydrodynamics, *J. Comp. Phys.*, 154:284–309.

Pulkkinen, A., and M.. Engels (2005), The role of 3D geomagnetic induction in the determination of the ionospheric currents from the ground geomagnetic data, *Ann. Geophys.*, 23, 909-917.

Pulkkinen, A., A. Viljanen, and R. Pirjola (2006), Estimation of geomagnetically induced current levels from different input data, *Space Weather*, Vol. 4, S08005, doi:10.1029/2006SW000229.

Pulkkinen, A., M. Hesse, S. Habib, L. Van der Zel, B. Damsky, F. Policelli, D. Fugate (2009), and W. Jacobs, Solar Shield: forecasting and mitigating space weather effects on high-voltage power transmission systems, *Natural Hazards*, doi:10.1007/s11069-009-

9432-x.

Pulkkinen, A., L. Rastätter, M. Kuznetsova, M. Hesse, A. Ridley, J. Raeder, H.J. Singer, and A. Chulaki (2010), Systematic evaluation of ground and geostationary magnetic field predictions generated by global magnetohydrodynamic models, *Journal of Geophysical Research*, 115, A03206, doi:10.1029/2009JA014537.

Pulkkinen, A., M. Kuznetsova, A. Ridley, J. Raeder, A. Vapirev, D. Weimer, R. S. Weigel, M. Wiltberger, G. Millward, L. Rastätter, M. Hesse, H. J. Singer and A. Chulaki (2011), Geospace Environment Modeling 2008-2009 Challenge: ground magnetic field perturbations, *Space Weather*, Vol. 9, S02004, doi:10.1029/2010SW000600.

Pulkkinen, A., L. Rastätter, M. Kuznetsova, H. Singer, C. Balch, D. Weimer, G. Toth, A. Ridley, T. Gombosi, M. Wiltberger, J. Raeder, R. Weigel(2011), Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations, *Space Weather*, Vol. 11, doi:10.1002/swe.20056.

Raeder, J. (2003), Global Magnetohydrodynamics A Tutorial, in *Space Plasma Simulation*, edited by J. Buchner, C. T. Dum, and M. Scholer, Springer Verlag, Berlin Heidelberg New York.

Raeder, J. (2006), Flux transfer events: 1. Generation mechanism for strong southward IMF, *Ann. Geophys.*, 24, 381.

Raeder, J., J. Berchem, and M. Ashour-Abdalla (1996), The importance of small scale processes in global MHD simulations: Some numerical experiments, in *The Physics of Space Plasmas*, edited by T. Chang and J. R. Jasperse, vol. 14, p. 403, MIT Cent. for Theoret. Geo/Cosmo Plasma Phys., Cambridge, Mass.

- Raeder, J., J. Berchem, and M. Ashour-Abdalla (1998), The Geospace Environment Modeling grand challenge: Results from a Global Geospace Circulation Model, *J. Geophys. Res.*, 103, 14,787.
- Raeder, J., Y. L. Wang, and T. Fuller-Rowell (2001a), Geomagnetic storm simulation with a coupled magnetosphere - ionosphere - thermosphere model, in *Space Weather, AGU Geophys. Monogr. Ser.*, edited by P. Song, G. Siscoe, and H. J. Singer, vol. 125, p. 377, American Geophysical Union.
- Raeder, J., Y. L. Wang, T. J. Fuller-Rowell, and H. J. Singer (2001b), Global simulation of space weather effects of the Bastille Day storm, *Sol. Phys.*, 204, 325.
- Raeder, J., D. Larson, W. Li, E. L. Kepko, and T. Fuller-Rowell (2008), OpenGGCM simulations for the THEMIS mission, *Space Sci. Rev.*, 141, 535, doi:10.1007/s11,21400894215.
- Raeder, J., P. Zhu, Y. Ge, and G. L. Siscoe (2010), OpenGGCM simulation of a substorm: Axial tail instability and ballooning mode preceding substorm onset, *J. Geophys. Res.*, 115, A00116.
- Raeder, J., et al. (2001c), Global simulation of the geospace environment modeling substorm challenge event, *J. Geophys. Res.*, 106, 381.
- Rastätter, L., M. Kuznetsova, A. Vapirev, A. Ridley, M. Wiltberger, A. Pulkkinen, M. Hesse and H.J. Singer (2011), Geospace Environment Modeling 2008-2009 Challenge: geosynchronous magnetic field, *Space Weather*, Vol. 9, S04005, doi:10.1029/2010SW000617.
- Rastätter, L. et al. (2013), Community-wide geospace model validation: computation of  $\Delta B$  at the CCMC, submitted to *Space Weather*.

- Richmond, A. D. (1995), Ionospheric electrodynamics using magnetic apex coordinates, *J. Geomag. Geoelectr.*, 47, 191.
- A. J. Ridley, K. C. Hansen, G. Tóth, D. L. De Zueew, T. I. Gombosi, and K. G. Powell (2002), University of Michigan MHD results of the GGCM metrics challenge, *J. Geophys. Res.*, 107(A10):1290.
- A. Ridley, T. Gombosi, and D. Dezeew (2004), Ionospheric control of the magnetosphere: conductance, *Annales Geophysicae*, 22:567–584.
- Robinson, R. M., R. R. Vondrak, K. Miller, T. Dabbs, and D. Hardy (1987), On calculating ionospheric conductances from the flux and energy of precipitating electrons, *J. Geophys. Res.*, 92, 2565.
- Roble, R. G., and E. C. Ridley (1994), A Thermosphere - Ionosphere - mesosphere - electrodynamics General Circulation Model (TIME-GCM): Equinox solar minimum simulations (30-500km), *Geophys. Res. Lett.*, 21, 417420.
- Skoug, R. M., Gosling, J. T., Steinberg, J. T., McComas, D. J., Smith, C. W., Ness, N. F., Hu, Q., Burlaga, L. F., (2004), Extremely high-speed solar wind: 2930 October 2003, *J. Geophys. Res.*, 109, A09102, doi:10.1029/2004JA010494.
- Siscoe, G., J. Raeder, and A. Ridley (2004), Transpolar potential saturation models compared, *J. Geophys. Res.*, 109, A09,203, doi:10.1029/2003JA010,318.
- Sokolov, I., E. V. Timofeev, Jun ichi Sakai, and Kazuyoshi Takayama (2002), Artificial wind – a new framework to construct simple and efficient upwind shock-capturing schemes, *J. Comput. Phys.*, 181:354–393.
- Sonnerup, B. U. O., and L. J. Cahill (1967), Magnetopause structure and attitude from Explorer 12 observations, *J. Geophys. Res.*, 72, 171.

- Sonnerup, B. U. O., and L. J. Cahill (1968), Explorer 12 observations of the magnetopause current layer, *J. Geophys. Res.*, 73, 1757.
- Tanskanen, E., A. Viljanen, T. Pulkkinen, R. Pirjola, L. Häkkinen, A. Pulkkinen, and O. Amm (2001), At substorm onset, 40% of AL comes from underground, *J. Geophys. Res.*, 106(A7), 13119-13134.
- Toffoletto, F., S. Sazykin, R. Spiro, and R. Wolf (2003), Inner magnetospheric modeling with the Rice Convection Model, *Space Sci. Rev.*, 107:175–196.
- Tóth, G., I. V. Sokolov, T. I. Gombosi, D. R. Chesney, C.R. Clauer, D. L. De Zeeuw, K. C. Hansen, K. J. Kane, W. B. Manchester, K. G. Powell, A. J. Ridley, I. I. Roussev, Q. F. Stout, O. Volberg, Richard A. Wolf, S. Sazykin, A. Chan, Bin Yu, and József Kóta (2005), Space Weather Modeling Framework: A new tool for the space science community, *J. Geophys. Res.*, 110:A12226.
- Tóth, G., D. L. De Zeeuw, T. I. Gombosi, and K. G. Powell (2006), A parallel explicit/implicit time stepping scheme on block-adaptive grids, *J. Comput. Phys.*, 217:722–758.
- Tóth, G., D. L. De Zeeuw, T. I. Gombosi, W. B. Manchester, A. J. Ridley, I. V. Sokolov, and I. I. Roussev (2007), Sun to thermosphere simulation of the October 28-30, 2003 storm with the Space Weather Modeling Framework, *Space Weather Journal*, 5:S06003.
- Tóth, G. et al (2012), Adaptive numerical algorithms in space weather modeling, *J. Comput. Phys.*, 231.
- VanZandt, T. E., W. L. Clark, and J. M. Warnock (1972), Magnetic apex coordinates: A magnetic coordinate system for the ionospheric f2 layer, *J. Geophys. Res.*, 77, 2406.

- Vennerstrom, S., T. Moretto, L. Rastaetter, and J. Raeder (2005), Field-aligned currents during northward interplanetary field: Morphology and causes, *J. Geophys. Res.*, 110, A06,205, doi:10.1029/2004JA010,802.
- Vennerstrom, S., T. Moretto, L. Rastaetter, and J. Raeder (2006), Modeling and analysis of solar wind generated contributions to the near-Earth magnetic field, *Earth, Planets, Space*, 58, 451-461.
- Viljanen, A., H. Nevanlinna, K. Pajunpää, and A. Pulkkinen (2001), Time derivative of the horizontal geomagnetic field as an activity indicator, *Ann. Geophys.*, 19, 1107-1118.
- Wang, W., M. Wiltberger, A. G. Burns, S. C. Solomon, T. L. Killeen, N. Maruyama, and J. G. Lyon (2004), Initial results from the coupled magnetosphere-ionosphere-thermosphere model: thermosphere-ionosphere responses, *Journal of Atmospheric and Solar-Terrestrial Physics*, 66(1), 1425-1441, doi:10.1016/j.jastp.2004.04.008.
- Weigel, R. S., A. J. Klimas, D. Vassiliadis (2003), Solar wind coupling to and predictability of ground magnetic fields and their time derivatives, *J. Geophys. Res.*, 108 (A7), 1298, doi:10.1029/2002JA009627.
- Weigel, R. S. (2007), Solar wind time history contribution to the day-of-year variation in geomagnetic activity, *J. Geophys. Res.*, 112, A10207, doi:10.1029/2007JA012324.
- Weimer, D. R., D. A. Gurnett, C. K. Goertz, J. D. Menietti, J. L. Burch, and M. Sugiura (1987), The current voltage relationship in auroral current sheets, *J. Geophys. Res.*, 92, 187.
- Weimer, D. R., and J. H. King (2008), Improved calculations of interplanetary magnetic field phase front angles and propagation time delays, *J. Geophys. Res.*, 113, A01105, doi:10.1029/2007JA012452.

- Weimer, D. R., C. R. Clauer, M. J. Engebretson, T. L. Hansen, H. Gleisner, I. Mann, and K. Yumoto (2010), Statistical maps of geomagnetic variations as a function of the interplanetary magnetic field, *J. Geophys. Res.*, 115, A10320, doi:10.1029/2010JA015540.
- Weimer, D. R. (2013), An empirical model of ground-level geomagnetic perturbations, *Space Weather*, in press, doi: 2012SW000880.
- D. T. Welling and A. J. Ridley (2010), Validation of swmf magnetic field and plasma, *Space Weather*, 8.
- Wintoft, P. (2005), Study of the solar wind coupling to the time difference horizontal geomagnetic field, *Annales Geophysicae*, 23, 1949-1957.
- Wiltberger, M., T. I. Pulkkinen, J. G. Lyon, and C. C. Goodrich (2000), MHD simulation of the magnetotail during the December 10, 1996, substorm, *Journal of Geophysical Research*, 105(A), 2764927664, doi:10.1029/1999JA000251.
- Wiltberger, M., J. G. Lyon, and C. C. Goodrich (2003), Results from the Lyon-Fedder-Mobarry global magnetospheric model for the electrojet challenge, *Journal of Atmospheric and Solar-Terrestrial Physics*, 65(1), 1213-1222, doi:10.1016/j.jastp.2003.08.003.
- Wiltberger, M., W. Wang, A. G. Burns, S. C. Solomon, J. G. Lyon, and C. C. Goodrich (2004), Initial results from the coupled magnetosphere ionosphere thermosphere model: magnetospheric and ionospheric responses, *Journal of Atmospheric and Solar-Terrestrial Physics*, 66(1), 1411-1423, doi:10.1016/j.jastp.2004.03.026.
- Wiltberger, M., R. S. Weigel, W. Lotko, and J. A. Fedder (2009), Modeling seasonal variations of auroral particle precipitation in a global-scale magnetosphere-ionosphere simulation, *Journal of Geophysical Research*, 114(A), 01204, doi:10.1029/2008JA013108.

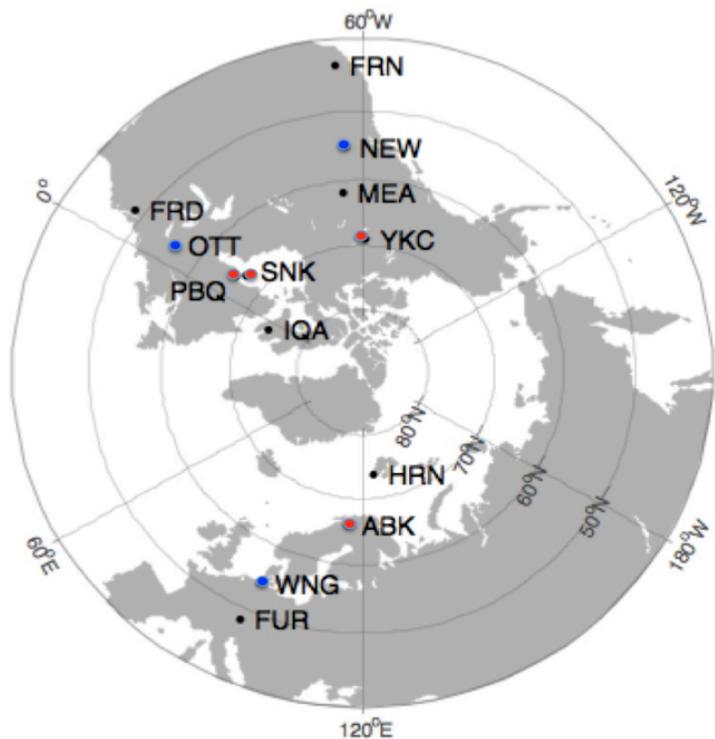
- Wiltberger, M., L. Qian, C.-L. Huang, W. Wang, R. E. Lopez, A. G. Burns, S. C. Solomon, Y. Deng, and Y. Huang (2012), CMIT study of CR2060 and 2068 comparing L1 and MAS solar wind drivers, *Journal of Atmospheric and Solar-Terrestrial Physics*, doi:10.1016/j.jastp.2012.01.005.
- Wolf, R.A., M. Harel, R. W. Spiro, G.-H. Voigt, P. H. Reiff, and C. K. Chen (1982), Computer simulation of inner magnetospheric dynamics for the magnetic storm of July 29, 1977, *J. Geophys. Res.*, 87:5949-5962.
- Zaharia, S., V. K. Jordanova, D. Welling, and G. Toth (2010), Self-consistent inner magnetosphere simulation driven by a global mhd model, *J. Geophys. Res.*, 115:A12228.
- Zhang, J., M.W. Liemohn, D.L. De Zeeuw, J.E. Borovsky, A.J. Ridley, G. Tóth, S. Sazykin, M.F. Thomsen, J.U. Kozyra, T.I. Gombosi, , and R.A. Wolf (2007), Understanding storm-time ring current development through data-model comparisons of a moderate storm, *J. Geophys. Res.*, 112:A04208.
- Zhang, J. J., C. Wang, and B. B. Tang (2012), Modeling geomagnetically induced electric field and currents by combining a global MHD model with a local one-dimensional method, *Space Weather*, 10, S05005, doi:10.1029/2012SW000772.
- Yu, Y., and A.J. Ridley (2008), Validation of the Space Weather Modeling Framework using ground-based magnetometers, *Space Weather*, 6.

**Table 1.** Geospace events studied in the validation activity. The last two columns give the minimum Dst index and the maximum Kp index of the event, respectively.

| Event # | Date and time                                      | min(Dst) | max(Kp) |
|---------|--|----------|---------|
| 1       | October 29, 2003 06:00 UT - October 30, 06:00 UT   | -353 nT  | 9       |
| 2       | December 14, 2006 12:00 UT - December 16, 00:00 UT | -139 nT  | 8       |
| 3       | August 31, 2001 00:00 UT - September 1, 00:00 UT   | -40 nT   | 4       |
| 4       | August 31, 2005 10:00 UT - September 1, 12:00 UT   | -131 nT  | 7       |
| 5       | April 5, 2010 00:00 UT - April 6, 00:00 UT         | -73 nT   | 8-      |
| 6       | August 5, 2011 09:00 UT - Aug 6, 09:00 UT          | -113 nT  | 8-      |

**Table 2.** The locations of the geomagnetic observatories used in the study. Bold typeface stations indicate the six stations (stations PBQ and SNK are alternates, see the text for details) used in the final analyses discussed in Section 5.

| Station name               | Station code | Geomagnetic lat | Geomagnetic lon | Scaling Factor |
|----------------------------|--------------|-----------------|-----------------|----------------|
| <b>Yellowknife</b>         | <b>YKC</b>   | 68.9            | 299.4           | 3.0            |
| Meanook                    | MEA          | 61.6            | 306.2           |                |
| <b>Newport</b>             | <b>NEW</b>   | 54.9            | 304.7           | 1.4            |
| Fresno                     | FRN          | 43.5            | 305.3           |                |
| Iqaluit                    | IQA          | 74.0            | 5.2             |                |
| <b>Poste-de-la-Baleine</b> | <b>PBQ</b>   | 65.5            | 351.8           | 3.0            |
| <b>Sanikiluaq</b>          | <b>SNK</b>   | 66.4            | 356.1           | 3.0            |
| <b>Ottawa</b>              | <b>OTT</b>   | 55.6            | 355.3           | 1.5            |
| Fredericksburg             | FRD          | 48.4            | 353.4           |                |
| Hornsund                   | HRN          | 73.9            | 126.0           |                |
| <b>Abisko</b>              | <b>ABK</b>   | 66.1            | 114.7           | 3.0            |
| <b>Wingst</b>              | <b>WNG</b>   | 54.1            | 95.0            | 1.0            |
| Furstenfeldbruck           | FUR          | 48.4            | 94.6            |                |



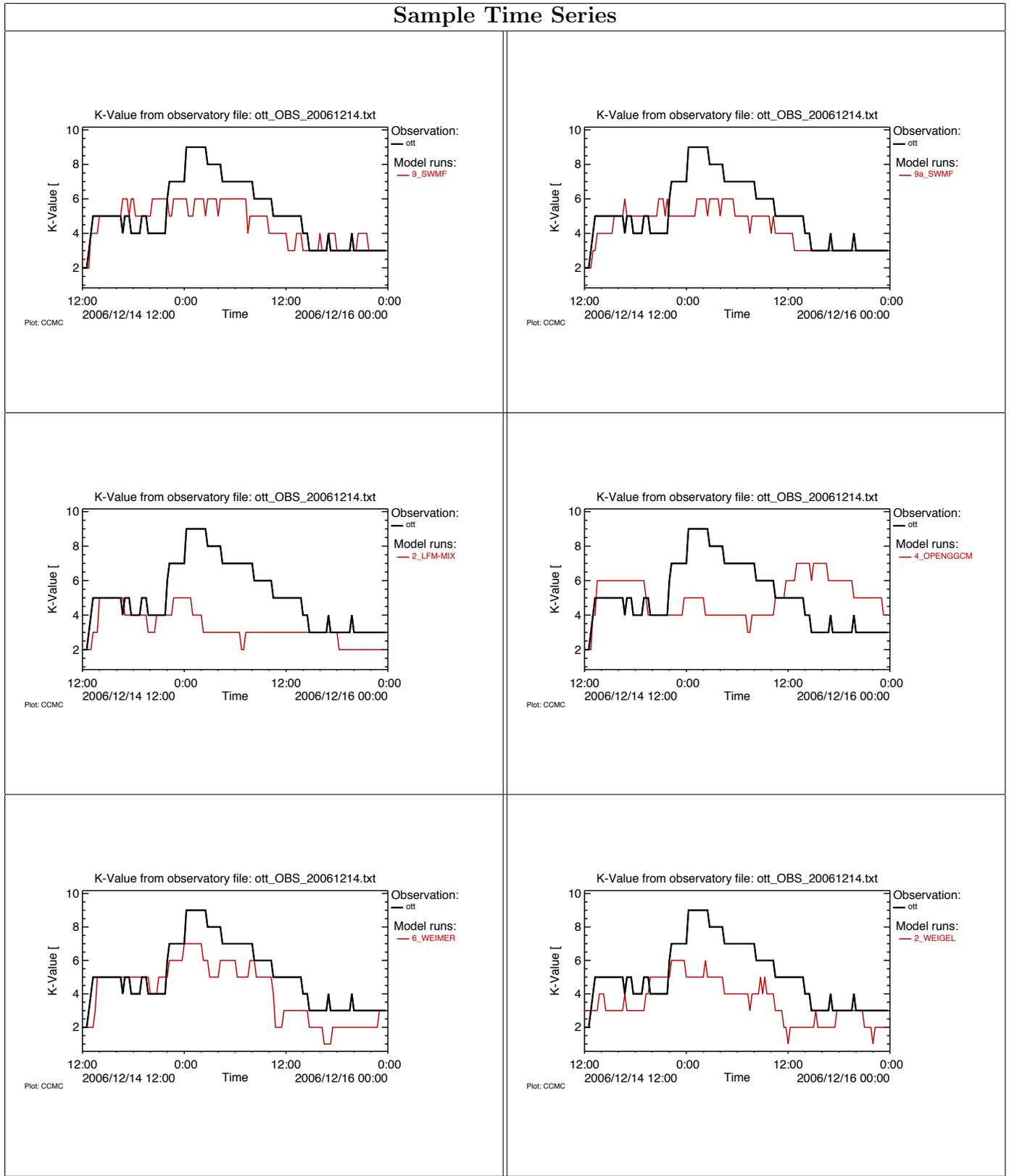
**Figure 1.** The locations and the station codes of the geomagnetic observatories used in the study. Geomagnetic dipole coordinates are used. Red and blue circles indicate high-latitude and mid-latitude stations, respectively, used in the final analyses in Section 5.

**Table 3.** Models analyzed in the validation effort. Each model is assigned a unique model identifier given by the leftmost column of the table. The table indicates the model setting, and if applicable, the number of cells and the minimum spatial resolution used in the global MHD part of the model. See text in Section 4 for details.

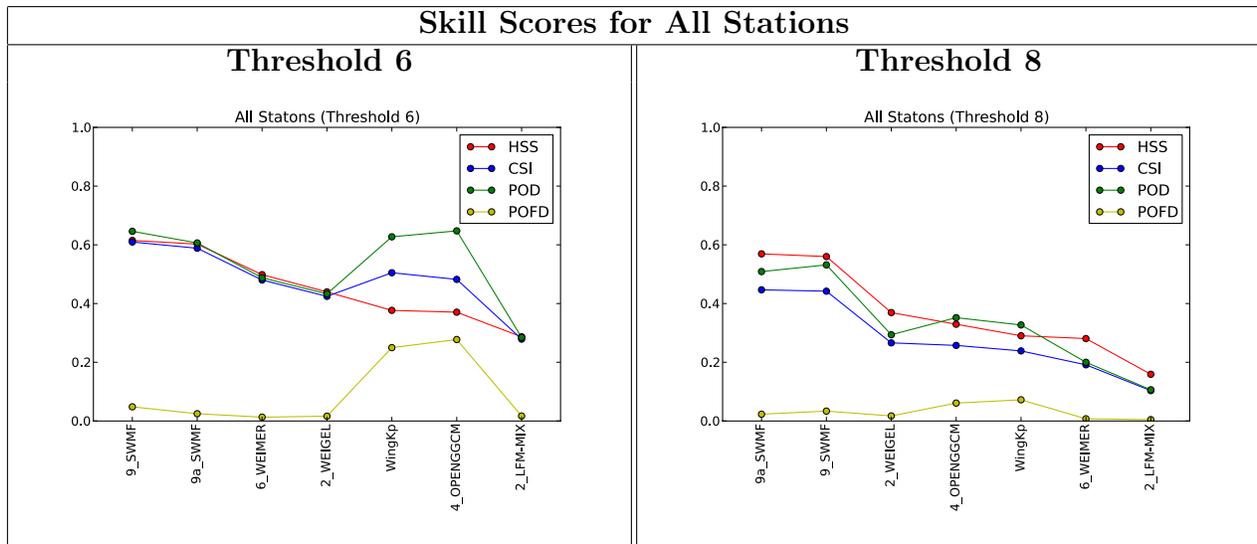
| Identifier | (model version) Model                                    | Grid (# of cells, min. re |
|------------|--|---------------------------|
| 2_LFM-MIX  | (LTR-2.1.1) LFM coupled with ionospheric electrodynamics | 163,000, $0.4 R_E$        |
| 3_WEIGEL   | empirical model  | N/A                       |
| 4_OPENGGCM | (OpenGGCM 4.0) global MHD coupled with CTIM              | 3.9 million, $0.25 R_E$   |
| 6_WEIMER   | empirical model  | N/A                       |
| 9_SWMF     | (SWMF 2011-01-31) BATS-R-US coupled with RIM and RCM     | 1 million, $0.25 R_E$     |

Acronyms:

|      |                                       |
|------|---------------------------------------|
| RIM  | Ridley Ionosphere Model               |
| RCM  | Rice Convection Model                 |
| CTIM | Coupled Thermosphere Ionosphere Model |



**Figure 2.** Time series of the observed (Black) and modeled (Red) k predictions for a particular mid-latitude station (OTT). Each panel shows a different model’s prediction.



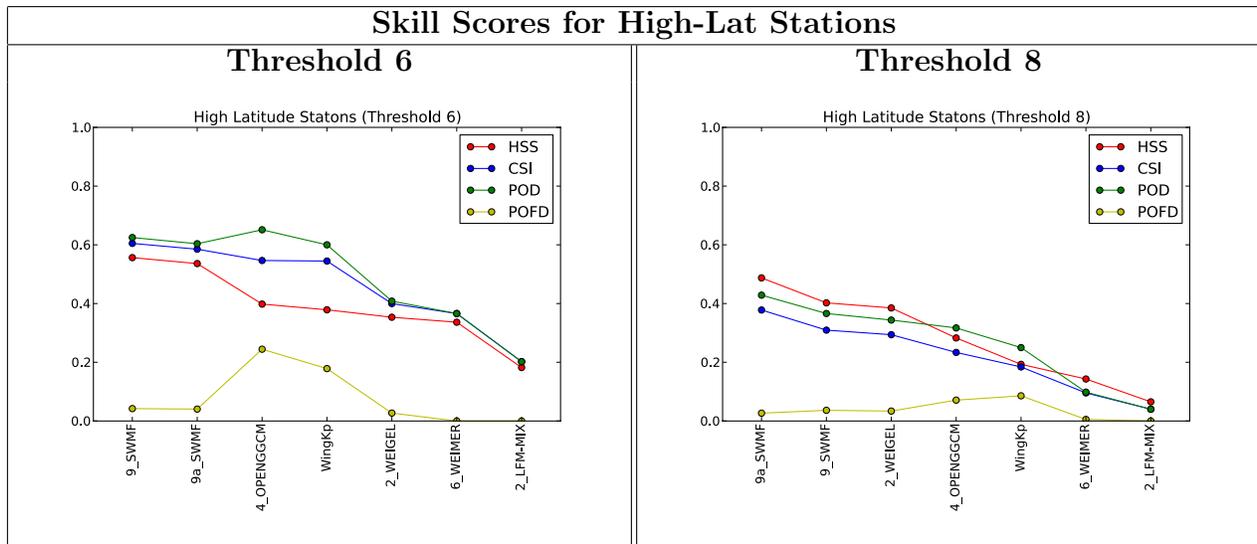
**Figure 3.** Heidke Skill Score (HSS) (red curve), Critical Success Index (CSI) (blue curve), Probability of Detection (POD) (green curve) and Probability of False Detection (POFD) (yellow curve) defined in Section 3 for the  $K$  thresholds 6 (left panel) and 8 (right panel). POD and POFD are obtained by integrating over the three mid-latitude stations and the three high-latitude stations. The models (see Table 3) are ordered according to their HSS. The model with the largest HSS is the leftmost in all panels.

| Run       | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|-----------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 9_SWMF    | 1240    | 1532      | 801   | 74      | 439    | 1458 | 0.61 | 0.61 | 0.65 | 0.05 |
| 9a_SWMF   | 1240    | 1532      | 752   | 38      | 488    | 1494 | 0.60 | 0.59 | 0.61 | 0.02 |
| 6_WEIMER  | 1240    | 1532      | 605   | 20      | 635    | 1512 | 0.50 | 0.48 | 0.49 | 0.01 |
| 2_WEIGEL  | 1240    | 1532      | 537   | 25      | 703    | 1507 | 0.44 | 0.42 | 0.43 | 0.02 |
| WingKp    | 1151    | 1117      | 722   | 279     | 429    | 838  | 0.38 | 0.50 | 0.63 | 0.25 |
| 4_OPENGGC | 1240    | 1532      | 803   | 425     | 437    | 1107 | 0.37 | 0.48 | 0.65 | 0.28 |
| 2_LFM-MIX | 1240    | 1532      | 353   | 26      | 887    | 1506 | 0.29 | 0.28 | 0.28 | 0.02 |

**Table 4.** Table for all stations, threshold 6

| Run       | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|-----------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 9a_SWMF   | 395     | 2377      | 201   | 55      | 194    | 2322 | 0.57 | 0.45 | 0.51 | 0.02 |
| 9_SWMF    | 395     | 2377      | 210   | 80      | 185    | 2297 | 0.56 | 0.44 | 0.53 | 0.03 |
| 2_WEIGEL  | 395     | 2377      | 116   | 41      | 279    | 2336 | 0.37 | 0.27 | 0.29 | 0.02 |
| 4_OPENGGC | 395     | 2377      | 139   | 145     | 256    | 2232 | 0.33 | 0.26 | 0.35 | 0.06 |
| WingKp    | 370     | 1898      | 121   | 137     | 249    | 1761 | 0.29 | 0.24 | 0.33 | 0.07 |
| 6_WEIMER  | 395     | 2377      | 79    | 18      | 316    | 2359 | 0.28 | 0.19 | 0.20 | 0.01 |
| 2_LFM-MIX | 395     | 2377      | 42    | 11      | 353    | 2366 | 0.16 | 0.10 | 0.11 | 0.00 |

**Table 5.** Table for all stations, threshold 8



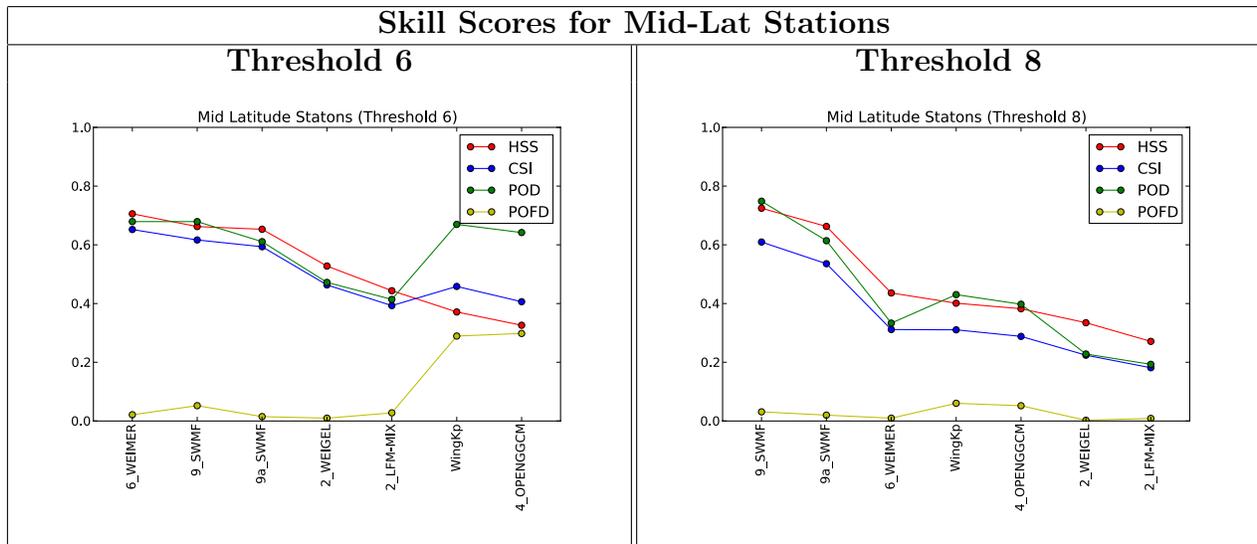
**Figure 4.** Heidke Skill Score (HSS) (red curve), Critical Success Index (CSI) (blue curve), Probability of Detection (POD) (green curve) and Probability of False Detection (POFD) (yellow curve) defined in Section 3 for the  $K$  thresholds 6 (left panel) and 8 (right panel). POD and POFD are obtained by integrating over the three high-latitude stations. The models (see Table 3) are ordered according to their HSS. The model with the largest HSS is the leftmost in all panels.

| Run       | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|-----------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 9_SWMF    | 757     | 593       | 473   | 25      | 284    | 568  | 0.56 | 0.60 | 0.62 | 0.04 |
| 9a_SWMF   | 757     | 593       | 457   | 24      | 300    | 569  | 0.54 | 0.59 | 0.60 | 0.04 |
| 4_OPENGGC | 757     | 593       | 493   | 145     | 264    | 448  | 0.40 | 0.55 | 0.65 | 0.24 |
| WingKp    | 700     | 398       | 420   | 71      | 280    | 327  | 0.38 | 0.54 | 0.60 | 0.18 |
| 2_WEIGEL  | 757     | 593       | 309   | 16      | 448    | 577  | 0.35 | 0.40 | 0.41 | 0.03 |
| 6_WEIMER  | 757     | 593       | 277   | 0       | 480    | 593  | 0.34 | 0.37 | 0.37 | 0.00 |
| 2_LFM-MIX | 757     | 593       | 153   | 0       | 604    | 593  | 0.18 | 0.20 | 0.20 | 0.00 |

**Table 6.** Table for high-latitude stations, threshold 6

| Run       | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|-----------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 9a_SWMF   | 224     | 1126      | 96    | 30      | 128    | 1096 | 0.49 | 0.38 | 0.43 | 0.03 |
| 9_SWMF    | 224     | 1126      | 82    | 41      | 142    | 1085 | 0.40 | 0.31 | 0.37 | 0.04 |
| 2_WEIGEL  | 224     | 1126      | 77    | 38      | 147    | 1088 | 0.39 | 0.29 | 0.34 | 0.03 |
| 4_OPENGGC | 224     | 1126      | 71    | 80      | 153    | 1046 | 0.28 | 0.23 | 0.32 | 0.07 |
| WingKp    | 212     | 886       | 53    | 76      | 159    | 810  | 0.19 | 0.18 | 0.25 | 0.09 |
| 6_WEIMER  | 224     | 1126      | 22    | 6       | 202    | 1120 | 0.14 | 0.10 | 0.10 | 0.01 |
| 2_LFM-MIX | 224     | 1126      | 9     | 0       | 215    | 1126 | 0.07 | 0.04 | 0.04 | 0.00 |

**Table 7.** Table for high-latitude stations, threshold 8



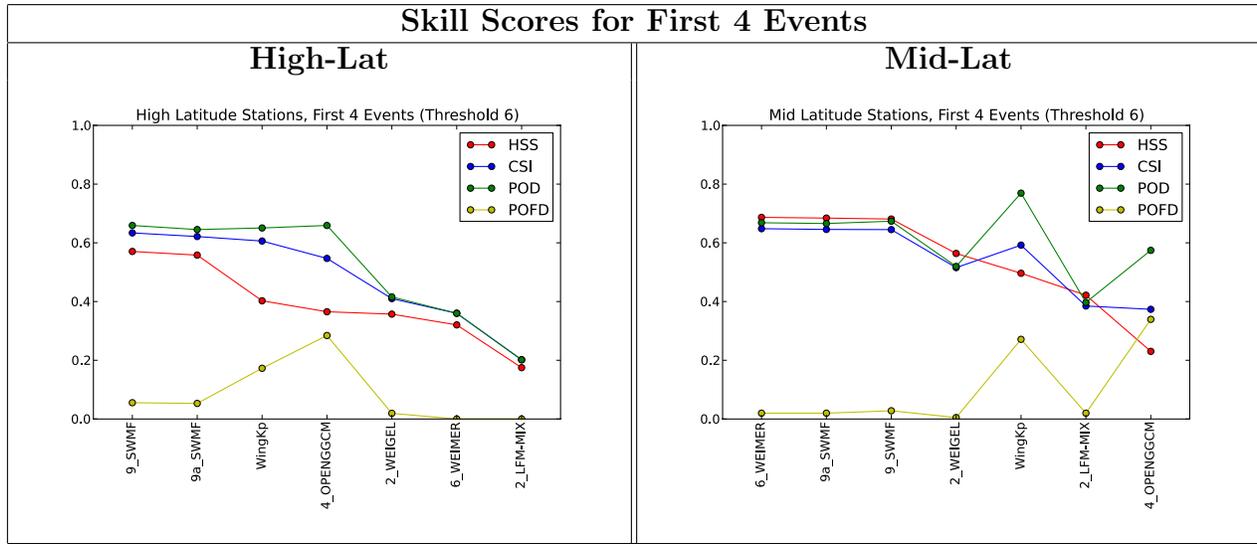
**Figure 5.** Heidke Skill Score (HSS) (red curve), Critical Success Index (CSI) (blue curve), Probability of Detection (POD) (green curve) and Probability of False Detection (POFD) (yellow curve) defined in Section 3 for the  $K$  thresholds 6 (left panel) and 8 (right panel). POD and POFD are obtained by integrating over the three mid-latitude stations. The models (see Table 3) are ordered according to their HSS. The model with the largest HSS is the leftmost in all panels.

| Run       | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|-----------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 6_WEIMER  | 483     | 939       | 328   | 20      | 155    | 919  | 0.71 | 0.65 | 0.68 | 0.02 |
| 9_SWMF    | 483     | 939       | 328   | 49      | 155    | 890  | 0.66 | 0.62 | 0.68 | 0.05 |
| 9a_SWMF   | 483     | 939       | 295   | 14      | 188    | 925  | 0.65 | 0.59 | 0.61 | 0.01 |
| 2_WEIGEL  | 483     | 939       | 228   | 9       | 255    | 930  | 0.53 | 0.46 | 0.47 | 0.01 |
| 2_LFM-MIX | 483     | 939       | 200   | 26      | 283    | 913  | 0.44 | 0.39 | 0.41 | 0.03 |
| WingKp    | 451     | 719       | 302   | 208     | 149    | 511  | 0.37 | 0.46 | 0.67 | 0.29 |
| 4_OPENGGC | 483     | 939       | 310   | 280     | 173    | 659  | 0.33 | 0.41 | 0.64 | 0.30 |

**Table 8.** Table for mid-latitude stations, threshold 6

| Run       | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|-----------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 9_SWMF    | 171     | 1251      | 128   | 39      | 43     | 1212 | 0.72 | 0.61 | 0.75 | 0.03 |
| 9a_SWMF   | 171     | 1251      | 105   | 25      | 66     | 1226 | 0.66 | 0.54 | 0.61 | 0.02 |
| 6_WEIMER  | 171     | 1251      | 57    | 12      | 114    | 1239 | 0.44 | 0.31 | 0.33 | 0.01 |
| WingKp    | 158     | 1012      | 68    | 61      | 90     | 951  | 0.40 | 0.31 | 0.43 | 0.06 |
| 4_OPENGGC | 171     | 1251      | 68    | 65      | 103    | 1186 | 0.38 | 0.29 | 0.40 | 0.05 |
| 2_WEIGEL  | 171     | 1251      | 39    | 3       | 132    | 1248 | 0.33 | 0.22 | 0.23 | 0.00 |
| 2_LFM-MIX | 171     | 1251      | 33    | 11      | 138    | 1240 | 0.27 | 0.18 | 0.19 | 0.01 |

**Table 9.** Table for mid-latitude stations, threshold 8



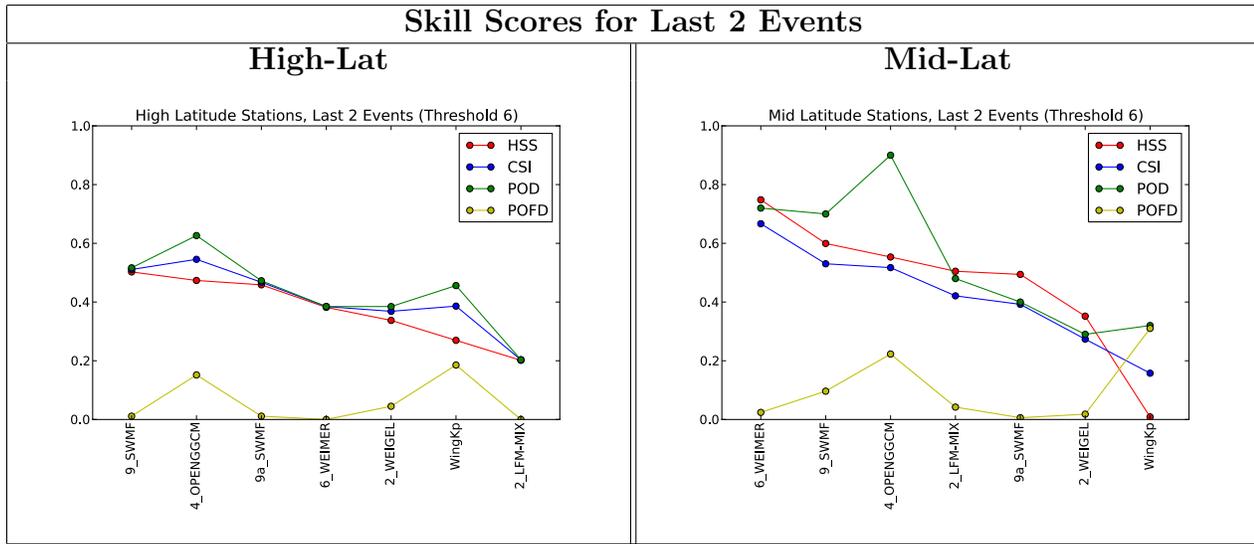
**Figure 6.** Heidke Skill Score (HSS) (red curve), Critical Success Index (CSI) (blue curve), Probability of Detection (POD) (green curve) and Probability of False Detection (POFD) (yellow curve) defined in Section 3 for  $K$  threshold of 6. High-Lat results (left panel) and Mid-Lat results (right panel). The models (see Table 3) are ordered according to their HSS. The model with the largest HSS is the leftmost in all panels.

| Run        | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|------------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 9_SWMF     | 575     | 415       | 379   | 23      | 196    | 392  | 0.57 | 0.63 | 0.66 | 0.06 |
| 9a_SWMF    | 575     | 415       | 371   | 22      | 204    | 393  | 0.56 | 0.62 | 0.65 | 0.05 |
| WingKp     | 518     | 220       | 337   | 38      | 181    | 182  | 0.40 | 0.61 | 0.65 | 0.17 |
| 4_OPENGGCM | 575     | 415       | 379   | 118     | 196    | 297  | 0.37 | 0.55 | 0.66 | 0.28 |
| 2_WEIGEL   | 575     | 415       | 239   | 8       | 336    | 407  | 0.36 | 0.41 | 0.42 | 0.02 |
| 6_WEIMER   | 575     | 415       | 207   | 0       | 368    | 415  | 0.32 | 0.36 | 0.36 | 0.00 |
| 2_LFM-MIX  | 575     | 415       | 116   | 0       | 459    | 415  | 0.17 | 0.20 | 0.20 | 0.00 |

**Table 10.** Table for first 4 events looking only at high-latitude stations, threshold 6

| Run       | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|-----------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 6_WEIMER  | 383     | 607       | 256   | 12      | 127    | 595  | 0.69 | 0.65 | 0.67 | 0.02 |
| 9a_SWMF   | 383     | 607       | 255   | 12      | 128    | 595  | 0.68 | 0.65 | 0.67 | 0.02 |
| 9_SWMF    | 383     | 607       | 258   | 17      | 125    | 590  | 0.68 | 0.65 | 0.67 | 0.03 |
| 2_WEIGEL  | 383     | 607       | 199   | 3       | 184    | 604  | 0.56 | 0.52 | 0.52 | 0.00 |
| WingKp    | 351     | 387       | 270   | 105     | 81     | 282  | 0.50 | 0.59 | 0.77 | 0.27 |
| 2_LFM-MIX | 383     | 607       | 152   | 12      | 231    | 595  | 0.42 | 0.38 | 0.40 | 0.02 |
| 4_OPENGGC | 383     | 607       | 220   | 206     | 163    | 401  | 0.23 | 0.37 | 0.57 | 0.34 |

**Table 11.** Table for first 4 events looking only at mid-latitude stations, threshold 8



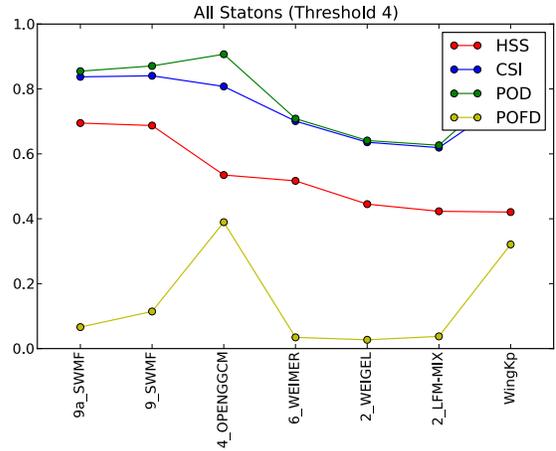
**Figure 7.** Heidke Skill Score (HSS) (red curve), Critical Success Index (CSI) (blue curve), Probability of Detection (POD) (green curve) and Probability of False Detection (POFD) (yellow curve) defined in Section 3 for  $K$  threshold of 6. High-Lat results (left panel) and Mid-Lat results (right panel). The models (see Table 3) are ordered according to their HSS. The model with the largest HSS is the leftmost in all panels.

| Run       | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|-----------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 9_SWMF    | 182     | 178       | 94    | 2       | 88     | 176  | 0.50 | 0.51 | 0.52 | 0.01 |
| 4_OPENGGC | 182     | 178       | 114   | 27      | 68     | 151  | 0.47 | 0.55 | 0.63 | 0.15 |
| 9a_SWMF   | 182     | 178       | 86    | 2       | 96     | 176  | 0.46 | 0.47 | 0.47 | 0.01 |
| 6_WEIMER  | 182     | 178       | 70    | 0       | 112    | 178  | 0.38 | 0.38 | 0.38 | 0.00 |
| 2_WEIGEL  | 182     | 178       | 70    | 8       | 112    | 170  | 0.34 | 0.37 | 0.38 | 0.04 |
| WingKp    | 182     | 178       | 83    | 33      | 99     | 145  | 0.27 | 0.39 | 0.46 | 0.19 |
| 2_LFM-MIX | 182     | 178       | 37    | 0       | 145    | 178  | 0.20 | 0.20 | 0.20 | 0.00 |

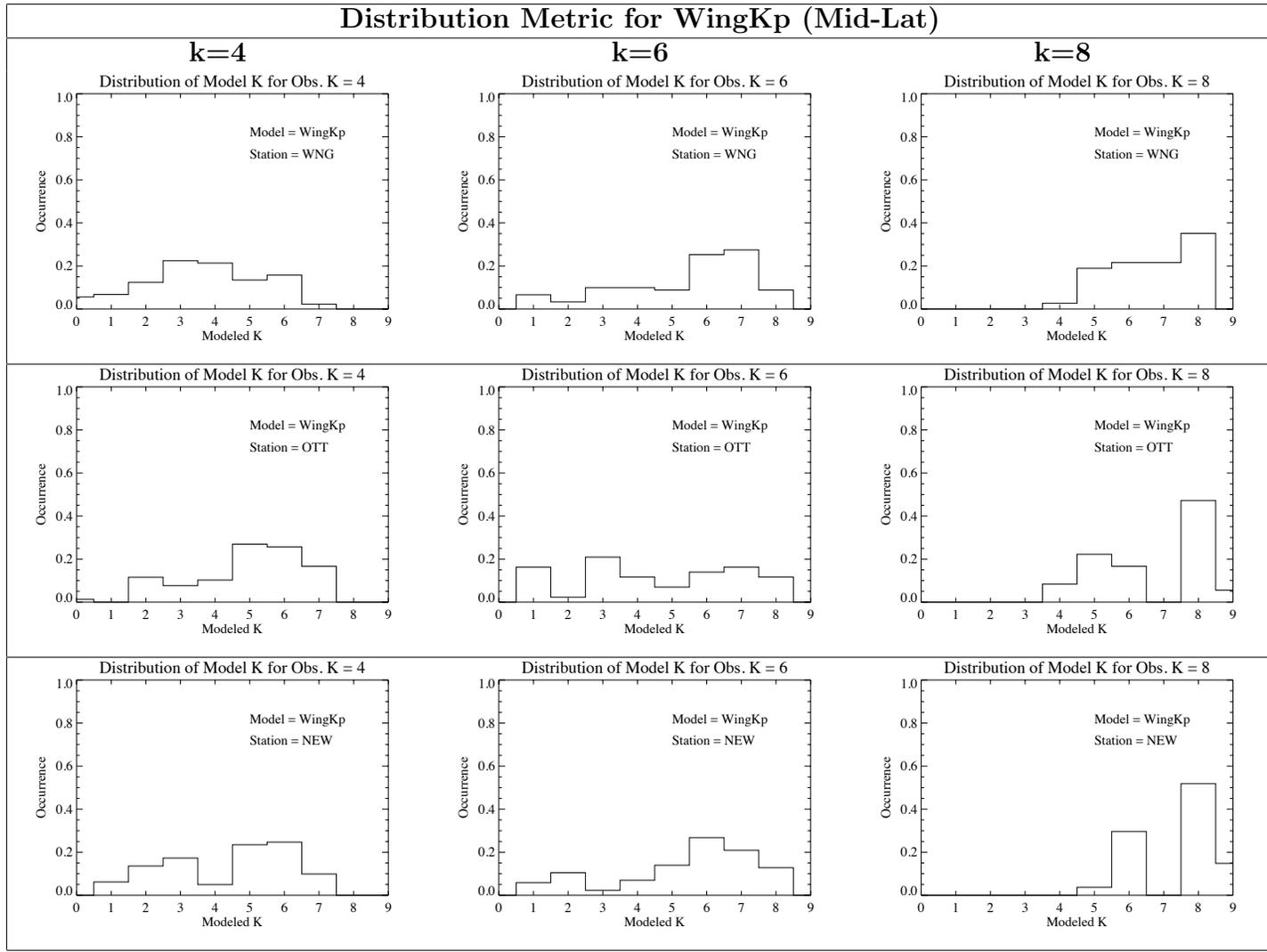
**Table 12.** Table for last 2 events looking at only high-latitude stations, threshold 6

| Run       | n_event | n_noevent | n_hit | n_false | n_miss | n_no | HSS  | CSI  | POD  | POFD |
|-----------|---------|-----------|-------|---------|--------|------|------|------|------|------|
| 6_WEIMER  | 100     | 332       | 72    | 8       | 28     | 324  | 0.75 | 0.67 | 0.72 | 0.02 |
| 9_SWMF    | 100     | 332       | 70    | 32      | 30     | 300  | 0.60 | 0.53 | 0.70 | 0.10 |
| 4_OPENGGC | 100     | 332       | 90    | 74      | 10     | 258  | 0.55 | 0.52 | 0.90 | 0.22 |
| 2_LFM-MIX | 100     | 332       | 48    | 14      | 52     | 318  | 0.50 | 0.42 | 0.48 | 0.04 |
| 9a_SWMF   | 100     | 332       | 40    | 2       | 60     | 330  | 0.49 | 0.39 | 0.40 | 0.01 |
| 2_WEIGEL  | 100     | 332       | 29    | 6       | 71     | 326  | 0.35 | 0.27 | 0.29 | 0.02 |
| WingKp    | 100     | 332       | 32    | 103     | 68     | 229  | 0.01 | 0.16 | 0.32 | 0.31 |

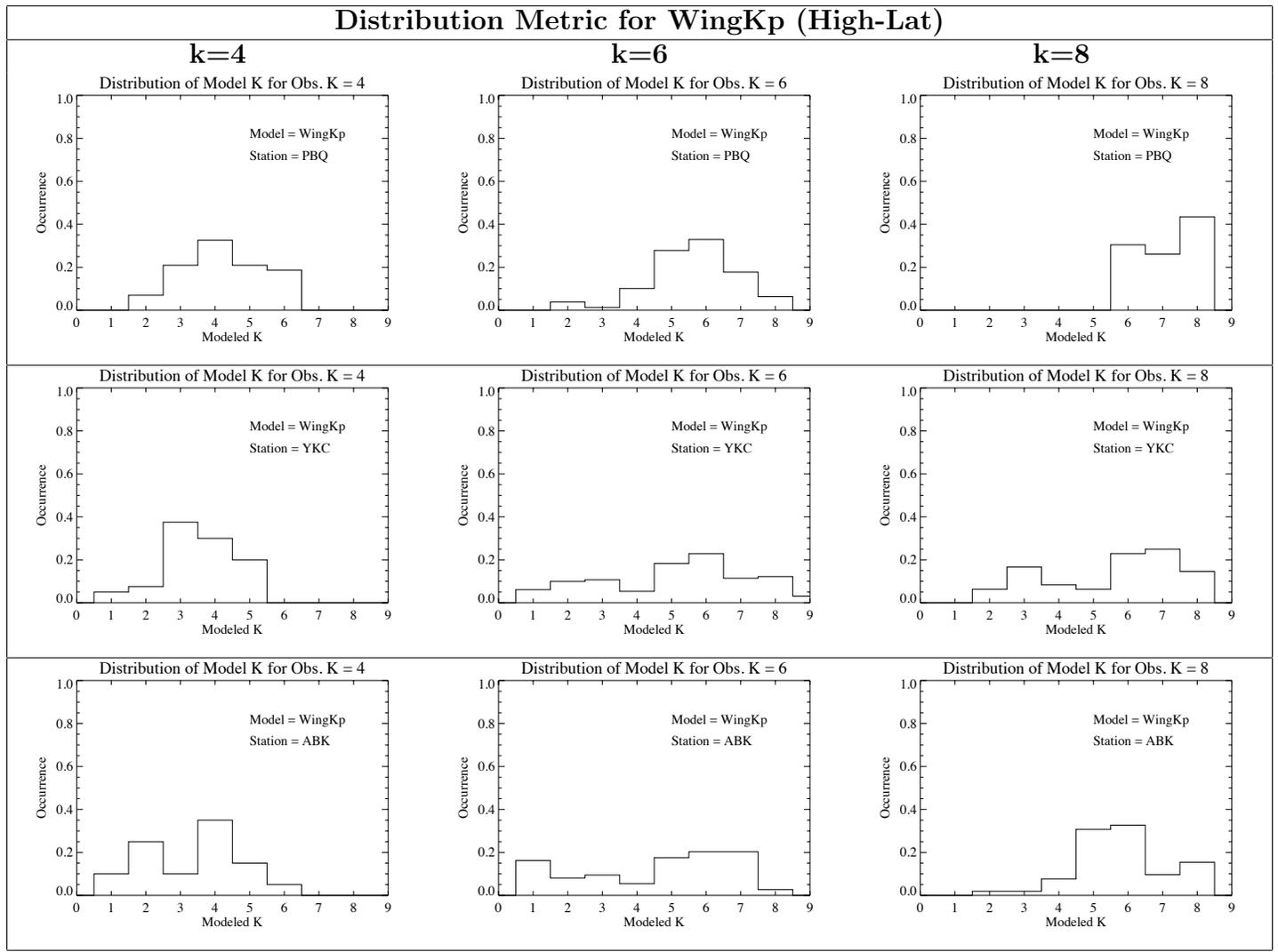
**Table 13.** Table for last 2 events looking at only mid-latitude stations, threshold 8



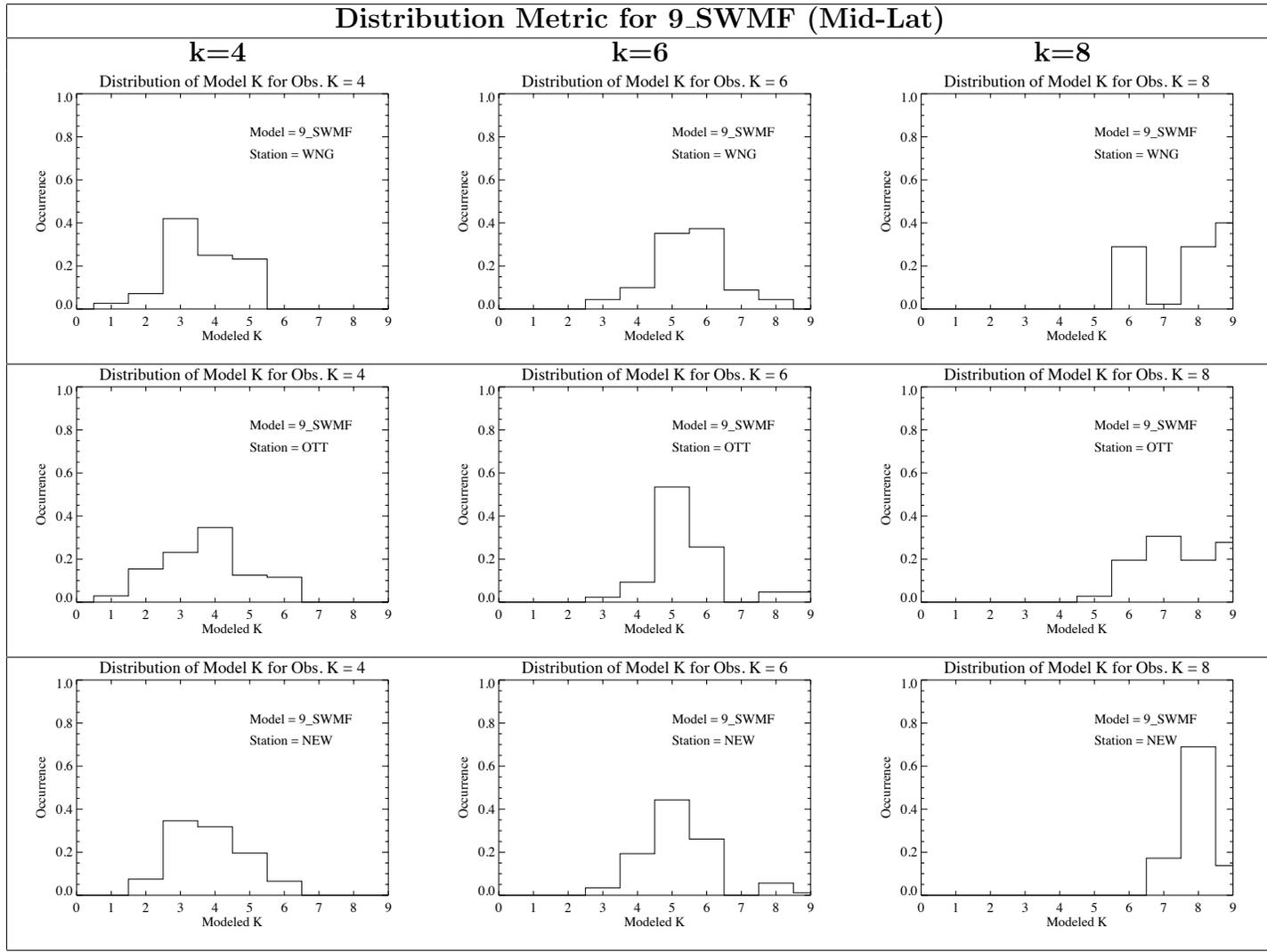
**Figure 8.** Heidke Skill Score (HSS) (red curve), Critical Success Index (CSI) (blue curve), Probability of Detection (POD) (green curve) and Probability of False Detection (POFD) (yellow curve) defined in Section 3 for the  $K$  threshold of 4. The panel shows POD and POFD obtained by integrating over all stations and events. The models (see Table 3) are ordered according to their HSS. The model with the largest HSS is the leftmost in all panels.



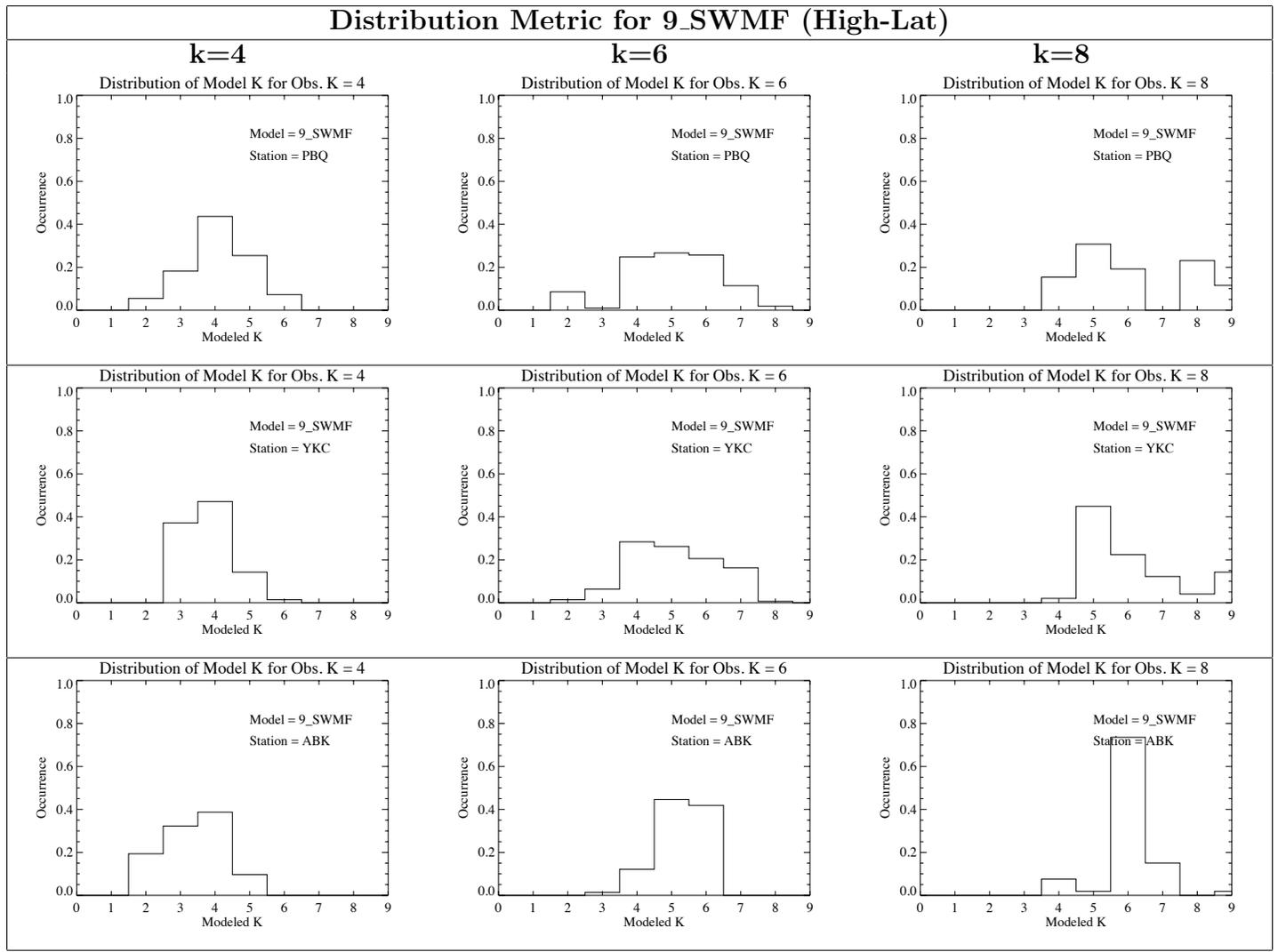
**Figure 9.** Distribution of WingKp Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different mid-latitude station.



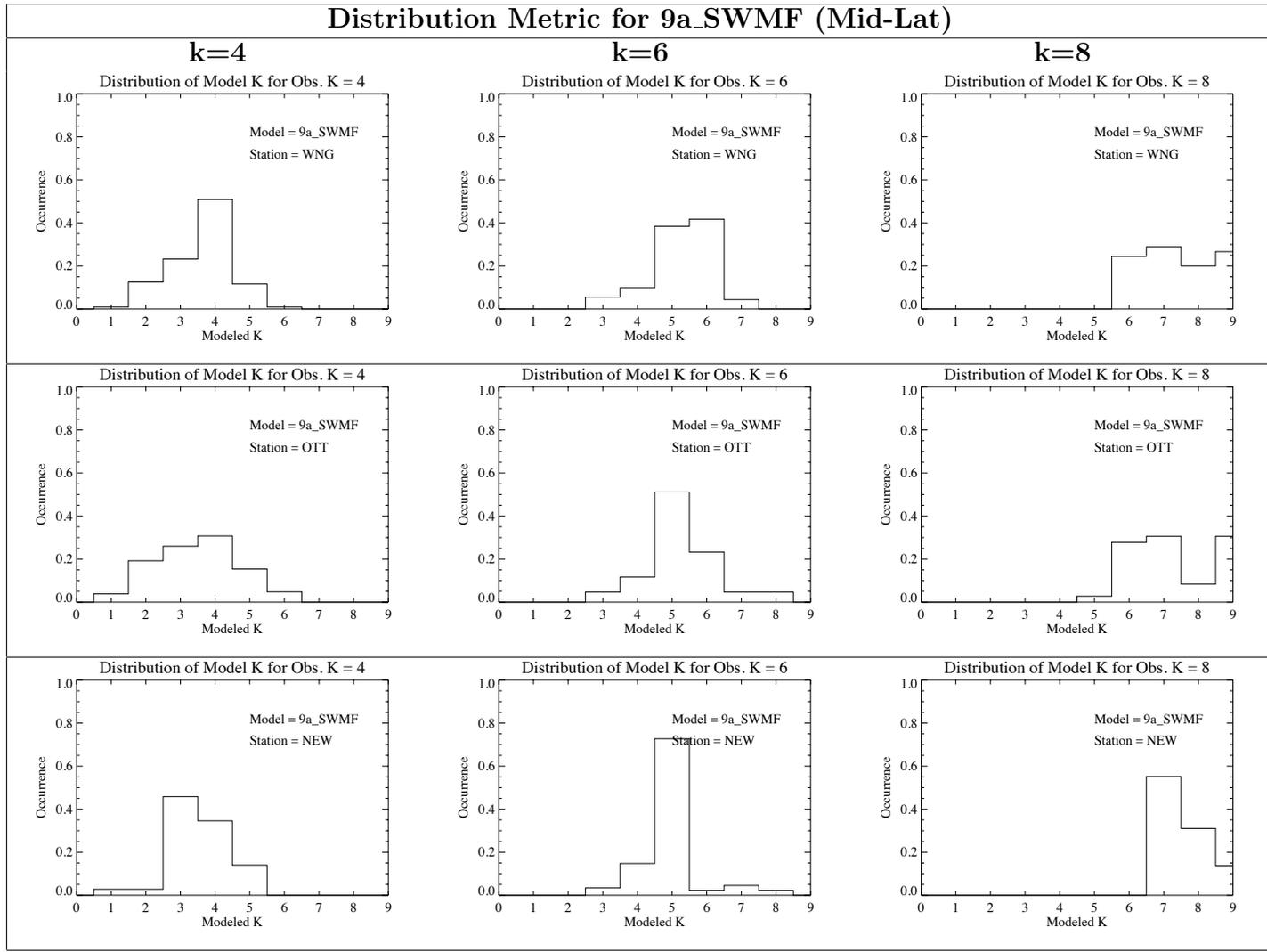
**Figure 10.** Distribution of WingKp Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different high-latitude station.



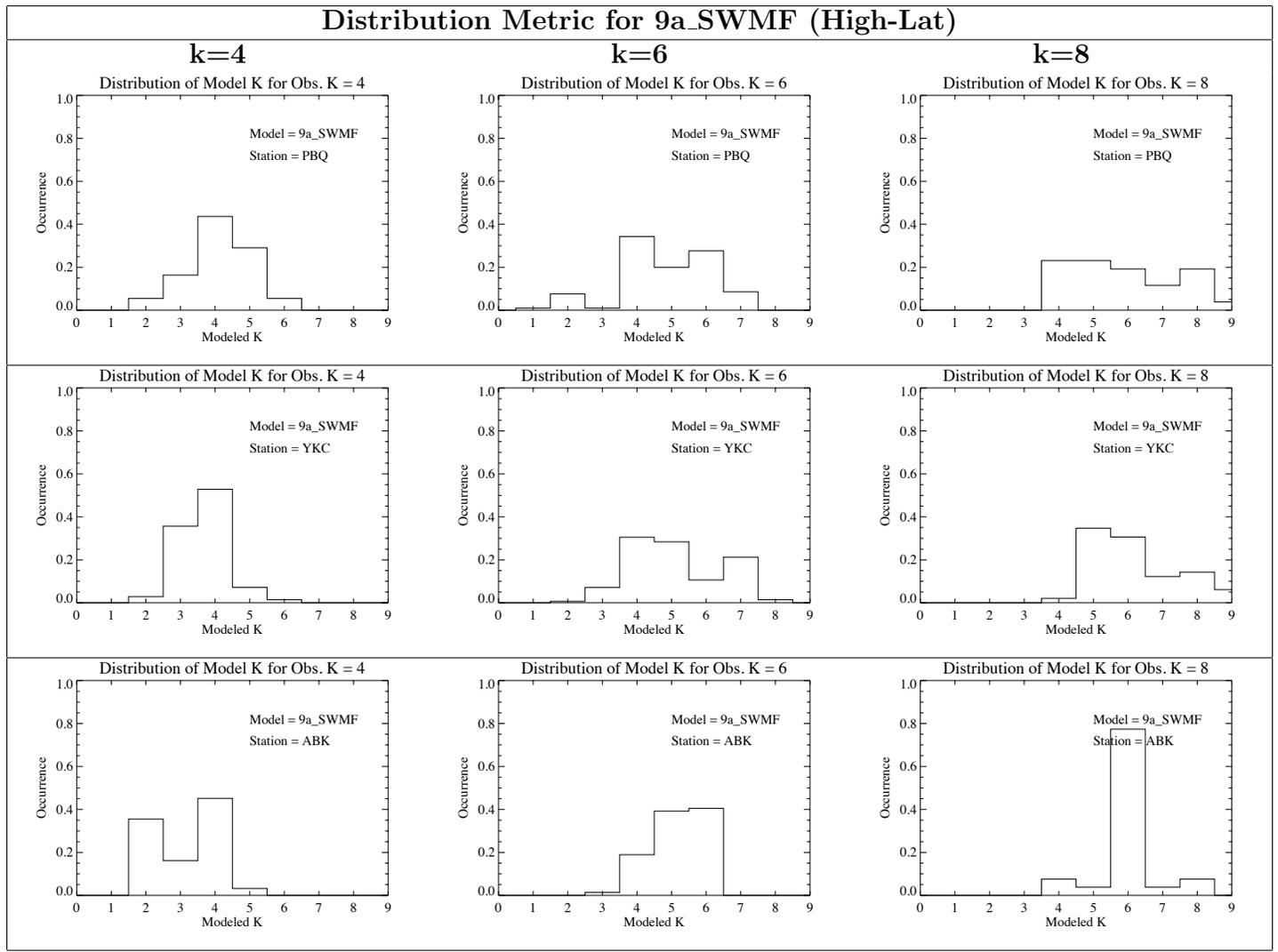
**Figure 11.** Distribution of 9\_SWMF Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different mid-latitude station.



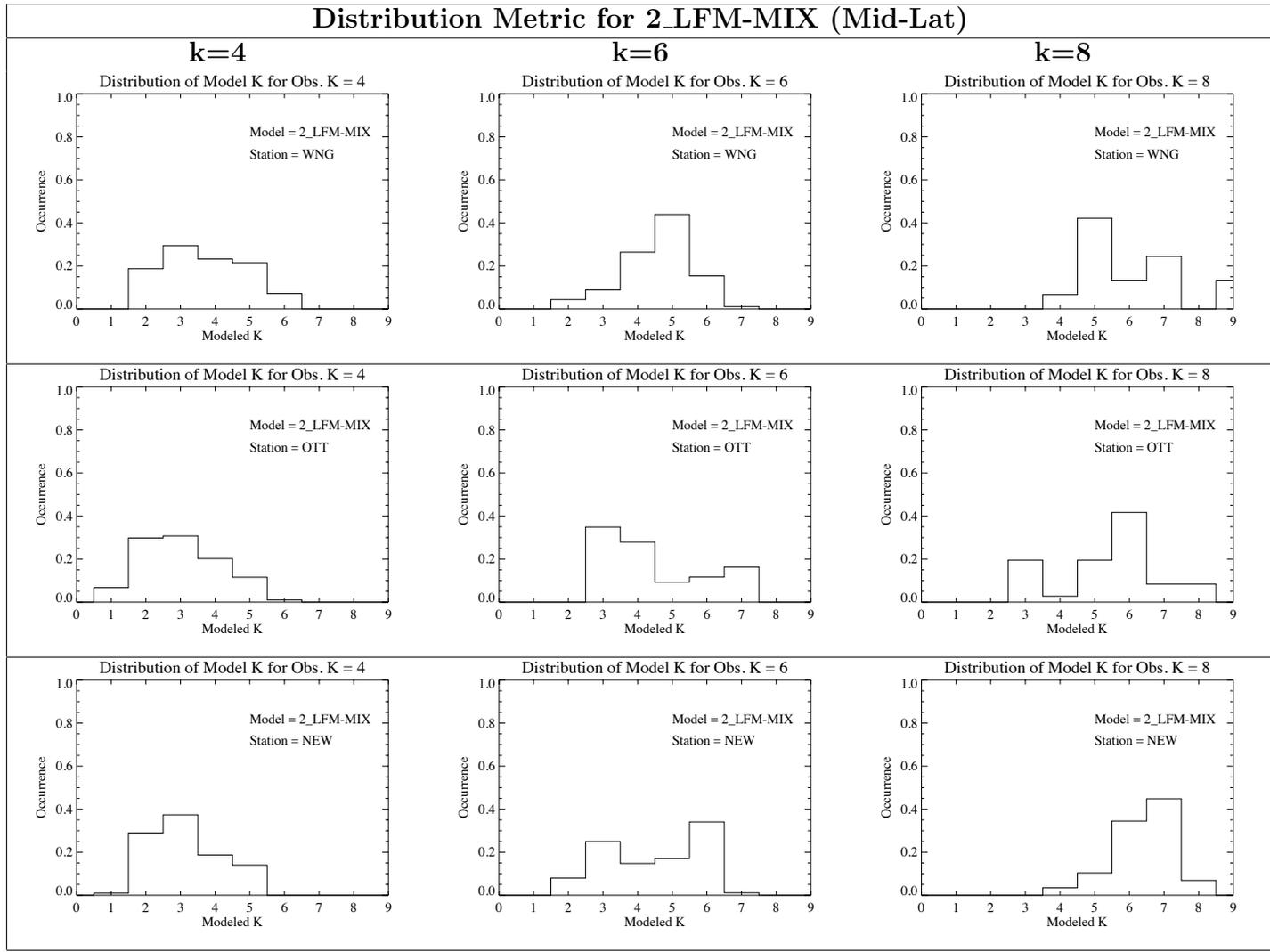
**Figure 12.** Distribution of 9\_SWMF Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different high-latitude station.



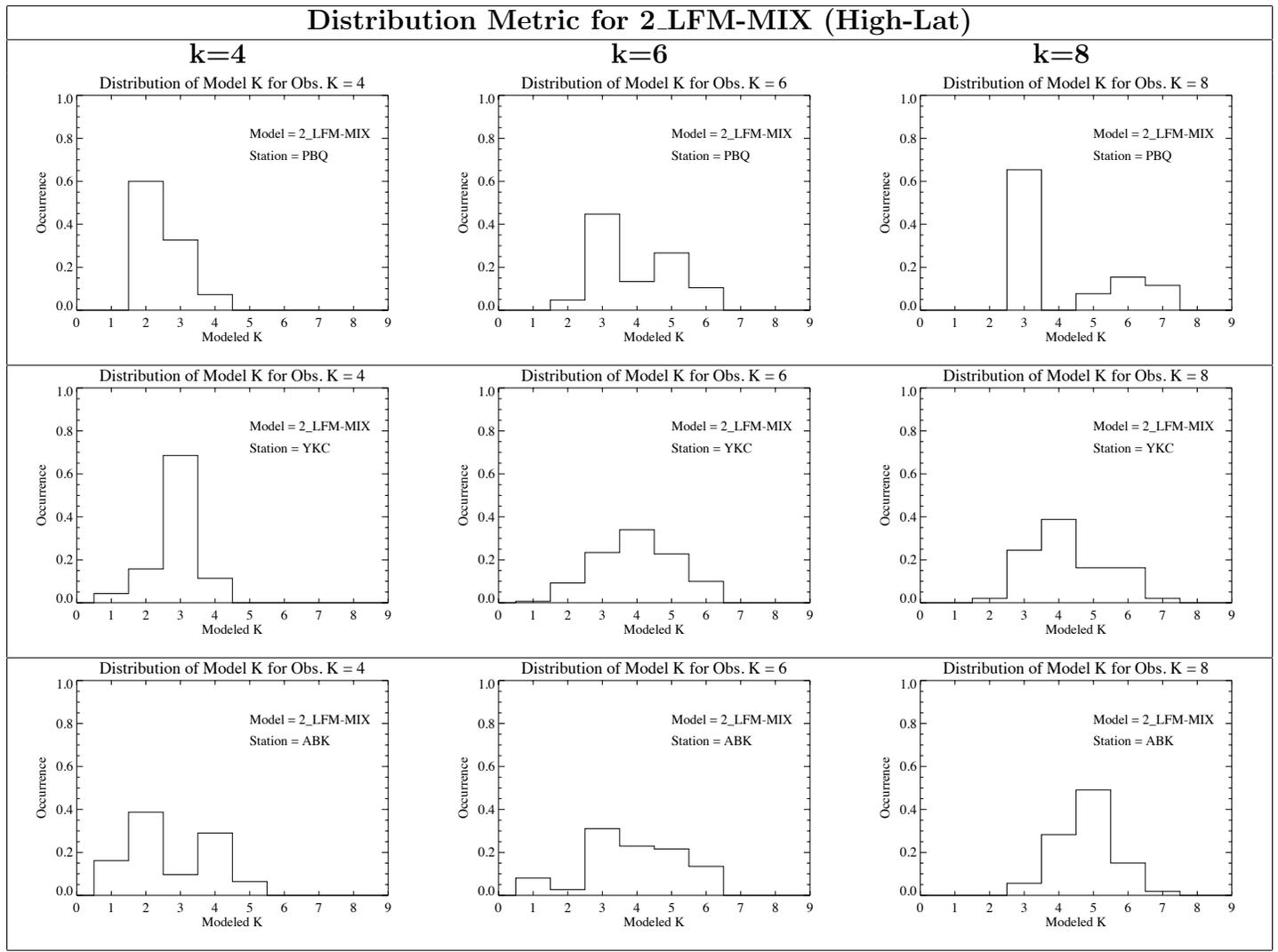
**Figure 13.** Distribution of 9a\_SWMF Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different mid-latitude station.



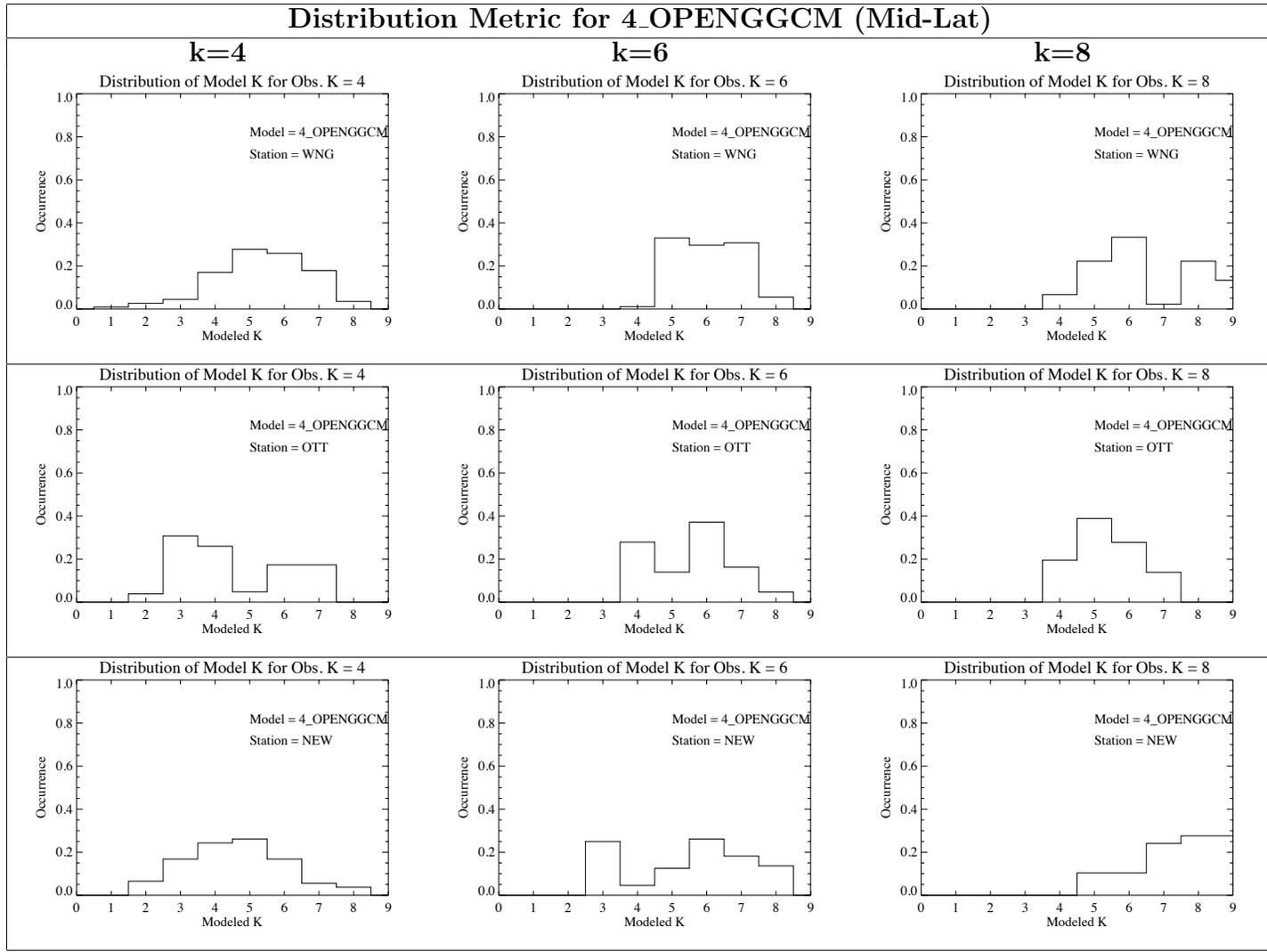
**Figure 14.** Distribution of 9a\_SWMF Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different high-latitude station.



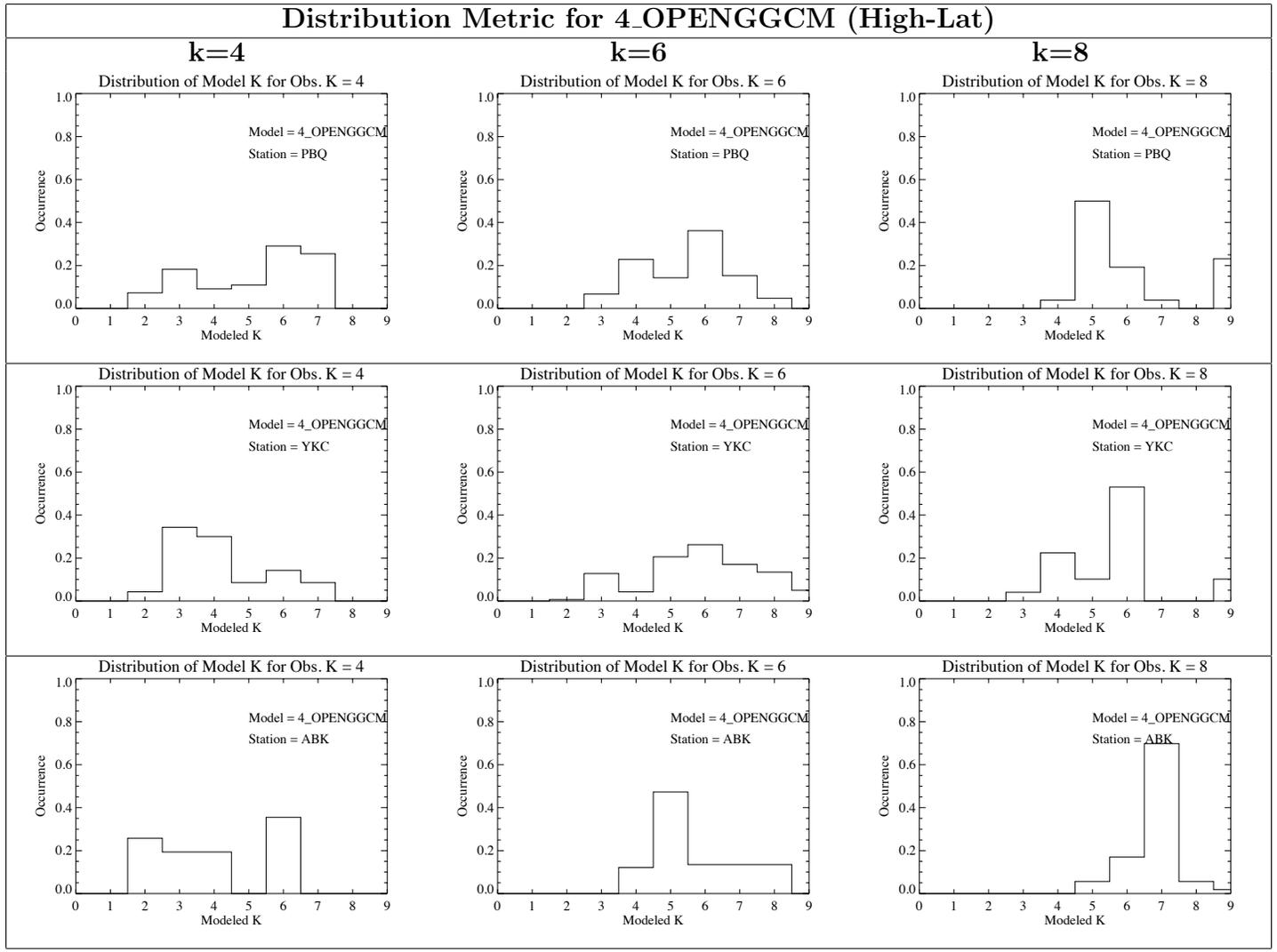
**Figure 15.** Distribution of 2\_LFM-MIX Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different mid-latitude station.



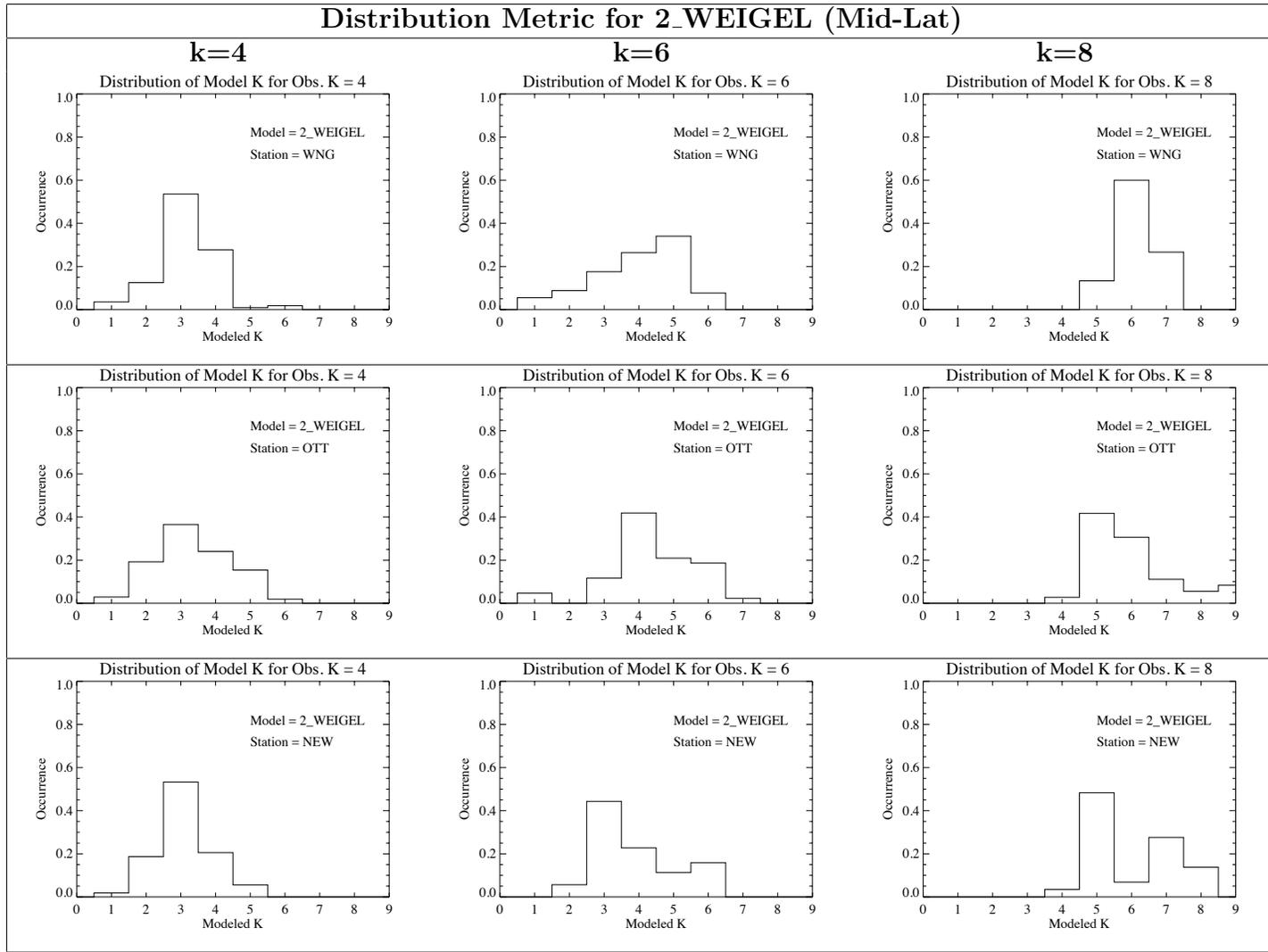
**Figure 16.** Distribution of 2\_LFM-MIX Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different high-latitude station.



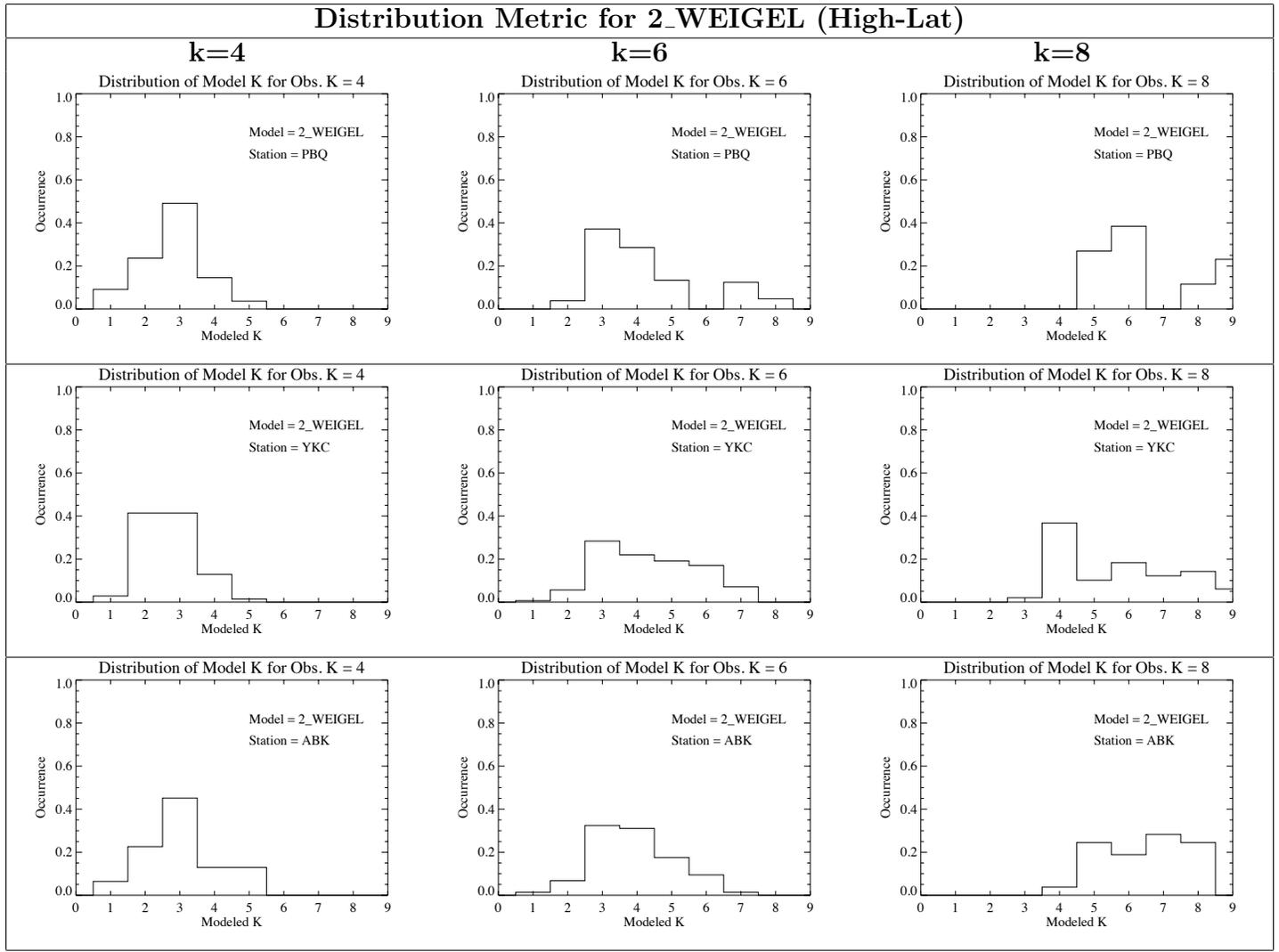
**Figure 17.** Distribution of 4\_OPENGGCM Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different mid-latitude station.



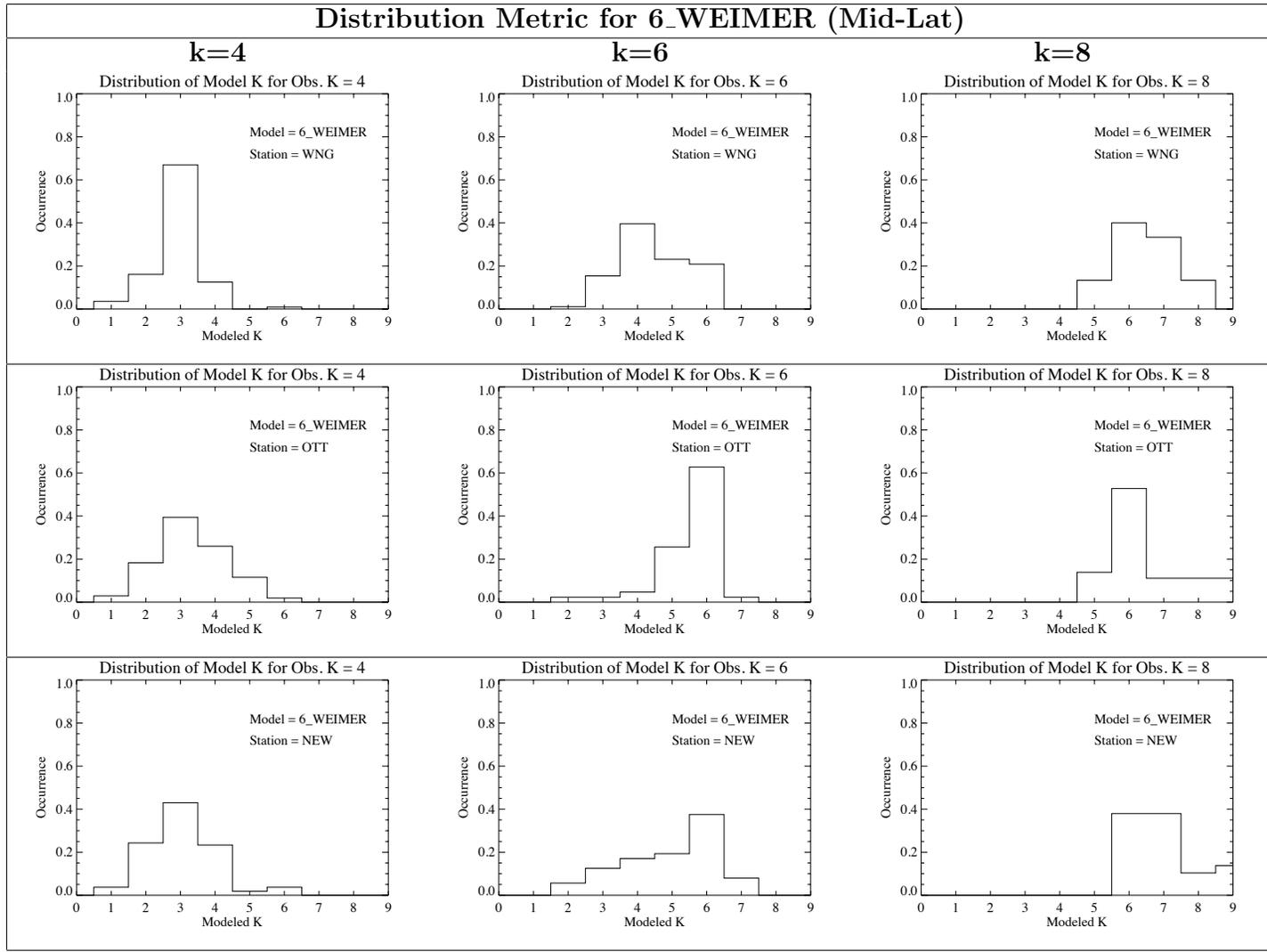
**Figure 18.** Distribution of 4\_OPENGGCM Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different high-latitude station.



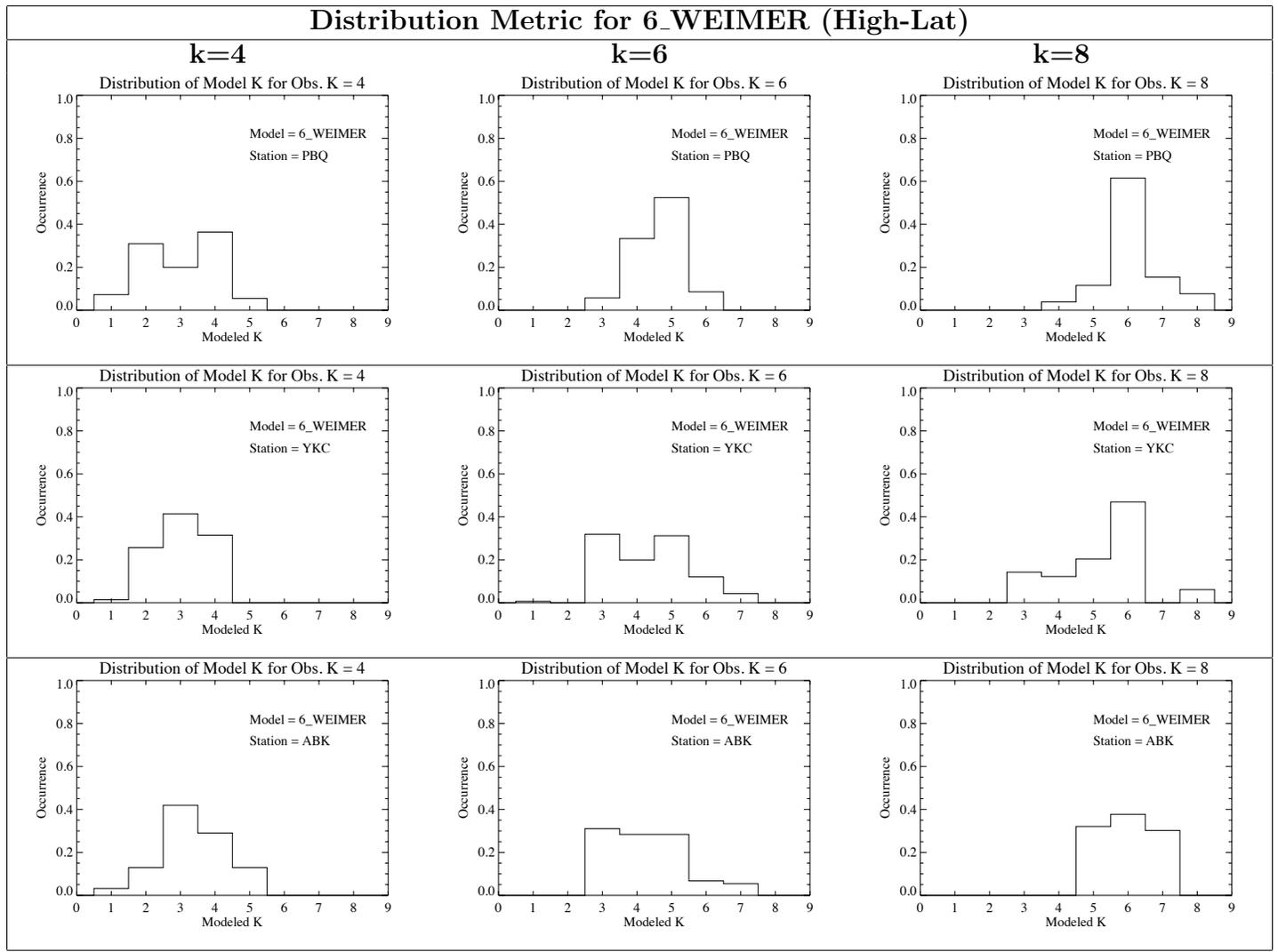
**Figure 19.** Distribution of 2\_WEIGEL Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different mid-latitude station.



**Figure 20.** Distribution of 2\_WEIGEL Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different high-latitude station.



**Figure 21.** Distribution of 6\_WEIMER Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different mid-latitude station.



**Figure 22.** Distribution of 6\_WEIMER Model predictions when  $K=4$  (left column),  $K=6$  (middle column), and  $K=8$  (right column). Each row presents results for a different high-latitude station.